User Guide: UDL Overlap Tool

This user guide is intended to provide an overview of the UDL Overlap tool and the associated data used by the tool.

The UDL Overlap tool supports the assessment of threatened and endangered species by summarizing the co-occurrence between potential pesticide use sites and species location. This tool uses GIS based datasets when considering co-occurrence and compiles results as a percent overlap (*i.e.,* the area of potential use and off-site movement found within the species location divided by the total area for the species). Generated using a number of publicly available land cover datasets, the EPA Use Data Layers (UDLs) delineate the potential pesticide use locations and areas where off-site movement could occur following application. The datasets delineating species range and designated critical habitat are provided by the U.S Fish and Wildlife Service (USFWS) and National Marine Fisheries Services (NMFS) of NOAA. This tool generates outputs for species ranges and designated critical habitats separately, with the resulting percent overlaps summarized by use and off-site distance. When assessing impacts to threatened and endangered species, the outputs from this tool are utilized to set the overlap category, as low (<5%), medium (<10% but >5%) or high (>10%).

Software Installation Instructions

For EPA Employees: ArcGIS Desktop/Pro and Python Installation Instructions

Create an online EPA GeoPlatform account. This will allow you to access online GIS resources maintained by Esri and the EPA. In addition to gaining access to the EPA GeoPlatform, users will need to request ArcGIS Desktop Software, which requires supervisor approval, a software installation ticket to EISD using the ServiceNow portal and contacting the EPA's National Geo-Spatial Support Team or NGST (more details below). The current version of ArcGIS Desktop Software is ArcPro, however, this tool will run with ArcMap 10.x with the installation the 64-bit background processing add-on.

To get started with GIS, email the EPA's Geospatial Support Team at GEOServices@epa.gov and request the installation of ArcGIS Desktop Software from EISD using the ServiceNow portal. Your message should indicate that you are a new GIS user in the agency and that you need to install ArcGIS Pro on your computer. Someone on the EPA's Geo-Spatial Support Team will provide support related to licensing, and the EISD IT staff will handle the software installations.

To create an EPA GeoPlatform account and request ArcGIS Desktop Software, follow the example below:

1. Using Chrome, go to https://epa.maps.arcgis.com
2. Click on "Sign In" (upper right corner)
3. Click on "EPA enterprise" (blue box)
4. If you are using a GFE laptop with PIV card, you will be signed in automatically. If not, you'll get the LAN ID sign in option.

5. Using ServiceNow open a Software Request ticket for ArcGIS Pro from the Service Catalog. (https://usepa.servicenowservices.com/)
6. Email the NGST and let them know you have submitted the ticket in ServiceNow so that they can support EISD with the configuration your GeoPlatform account with licensing for Pro. (GeoServices@epa.gov, CC Connolly.jennifer@epa.gov)

For Non-EPA Employees: ArcGIS Desktop/Pro and Python Installation Instructions

Work with your IT staff to have ArcPro or ArcMap with 64-bit background processing installed on your computer.

Integrated development environment (IDE)

The UDL Overlap Tool is written in Python language. Running this Python code requires a Python code editor referred to an Integrated Development Environment (IDE). An IDE is a software application that combines tools to write, test and execute Python code. Your ArcPro installation includes an IDE called integrated development and learning environment (IDLE), designed specifically to work with GIS data in Python. For this reason, IDLE is the recommended IDE for this tool. If you wish to use a different IDE, you can; however, additional configuration may be needed. Specifically, your preferred IDE will need to use the ArcGIS Python install to access the arcpy and pandas (1.2.3) modules.

You can open IDLE from the ArcGIS folder in the Windows "Start Menu" or by right clicking on a Python script and selecting "Edit with IDLE (ArcGIS Pro)" or IDLE 64-bit IDLE option (**Figure 1**). This tool requires the use of 64-bit processing for summarizing the datasets.
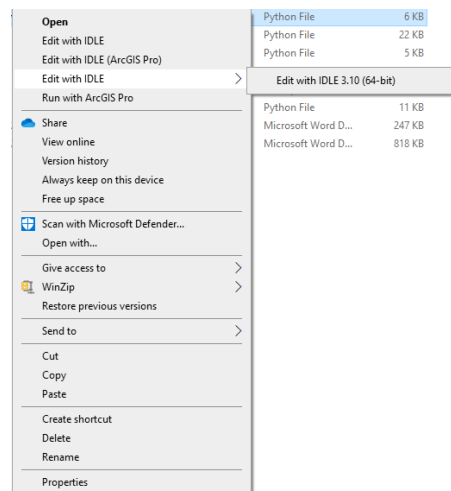

Figure 1. Opening IDLE by right clicking on a Python Script

Standard Inputs

At the start of your assessment, download the most up-to-date data and scripts that are available for download with the tool. The scripts and standard input files used by the UDL Overlap tool are routinely updated. As a result, for the purposes of versioning, each assessment will identify the dates of the data

sources so that they are reproducible. Specific data used in the assessments will be cited and summarized.

UDL Overlap Tool: Files to download:
- UDL_Overlap_Scripts.zip – Contains the Python Script
- Input.zip – Contains the input directory with all standard input files

After download, unzip these two files locally before use. Inputs should be prepared and organized before starting the overlap. You should save a copy of the scripts locally (i.e., your personal OneDrive, external hard drive, or C: drive). This will help with version control, safeguard the files, and ensure less runtime issues when running Python scripts.

The input data used by the UDL Overlap tool is found in the input.zip and is referred to the input directory. The Python scripts for the UDL Overlap Tool are found in the UDL_Overlap_Scripts.zip file. The only script that needs to be edited is the paths.py, see the next section for details. After updating the location of the input directory, and the assessment specific information in paths.py, each of the numbered scripts should be executed in numerical sequence.

Updating the Paths.py Script

General notes when editing the paths.py script:

- Any text following '#' is there for informational purposes only
- All text must be in single or double quotations
- Strings of text or numbers should be separated by a comma
- The "T" and "F" in True/False boolean statements need to be capitalized

Steps for updating the paths.py script:
1. Right click paths.py in Finder and click "Edit with IDLE" or "Edit with IDLE (ArcGIS Pro)".
   a. If you are using an IDE other than IDLE, then click "Edit with 'Your Desired IDE'"

General Parameters
2. Set the chemical name and the location of your input and output directories.
   a. chemical_name: should be the name of the chemical or product that you are working with (e.g., '2,4-D').
   b. input_dir: path to the location where you saved the inputs directory unzipped from the input.zip file
   c. output_dir: the location where you choose to save your outputs
   d. You can copy the path to these location from your File Explorer. An r must be present before your path occurs in double quotations. The r tells python which way the slashes in the path need to go and allows for spaces in the path. For example:
      i. r"C:\Users\Demo\input"
   e. overwrite_inter_data: will typically be set to True.
      i. When set to True new data tables will be generate each time you execute the tool

ii. This can be changed to False if you need to update just one use or variable and you do not wish to regenerate other work that was previously done.

iii. Reach out to the EFED GIS team with questions if you are unsure how to do this.

Chemical and Species-Specific Parameters

    a. Each step below outlines the edits based on the section of the paths.py script

    b. Each section starts with "# ### [Section Name]"

3. Range or Critical Habitat run
   a. range_run = True will run range files
   b. range_run = False will run critical habitat files
      i. There are two different types of GIS files that have been pre-processed in the input directory, one for range and one for critical habitat.
      ii. This parameter determines if you want to summarize the range or critical habitat.
      iii. You will need to run both range and critical habitat but can only run one at a time.
4. Taxa of concern section.
   a. You can remove taxa that are not of concern for a specific chemical. Be sure to consider both direct and indirect effects before removing taxa.
   b. taxa_concern = The taxa of concern will go between the two brackets, with each one in single or double quotations and separated by comma.
   c. USFWS and NMFS group taxa differently from EFED, so the in-line comments/text in this section provides guidance related to the differences in the groups for your reference.
   d. If you are unsure of which taxa are of concern or are still working to figure it out, you can run all taxa. Once you determine the taxa of concern, you can then re-run the tool or delete the taxa that aren't needed from the final output tables.
5. Regional restrictions
   a. Restricts the overlap to only those regions that are part of the registration based on the label. The regions include the conterminous United States (CONUS), non-conterminous states, and associated US territories.
   b. regions = is the list of all regions to be included.
      i. Example: regions = ['CONUS', 'AK']
6. State restrictions
   a. Restricts the overlap to specific states within the conterminous United States that are part of the registration based on the label.
   b. list_pol_ids = is the list of all states to be included. Always use the two-letter abbreviation for the state.
      i. Example: list_pol_ids = ['AR', 'FL', 'MS']
      ii. This output would only include species found within the above three states. If your label doesn't include a sub-set of states, leave the brackets as is.
7. Outputs that will be generated
   a. grouping_col is used to identify the types of output table you wish to generate. There three types of tables that can be generated:

      i.   Species: output table summarizes results across the whole species range or critical habitat by use. (Include EntityID, this is required)

      ii.   Species/state: output table summarizes results for the species range or critical habitat for each state it occurs in by use. (Include STUSPS, optional)

      iii.   Species state and county: output summarizes results for the species range or critical habitat for each state and county it occurs in by use. (Include GEOID, optional)

           1.   As a note, county tables take the longest to generate. It is suggested to start with the species and state overlap and generate county tables as needed.

  b.   Default parameter will generate the species and species/state output tables

      i.   grouping_col = ['EntityID', 'STUSPS']

      ii.   Add or remove variables to the parameter based on your needs

8.   Drift and runoff considerations

  a.   This section determines what distances (in meters) from the field will be included in your output tables for consideration of off-site movement. Note, distances are available in 30-meter increments (e.g., 0, 30, 60, etc.).

  b.   If specific distances for drift and/or runoff concerns are known for the chemical, those values can be added to the bins parameter.

  c.   If distances are unknown, or you are still determining this for your chemical, you can start with the default inputs of 0 (on-field), 30 (adjacent to field), 305 (distance limit for ground application in meters), 792 (distance for aerial application limit in meters) and 1,500 m (distance developed for runoff to wetlands).

      i.   The more distances you run, the longer the analysis will take.

  d.   Full_impact parameter will generally remain True.

      i.   Full impact means each off-site distance interval is inclusive of previous interval. For example, 305 is the limit of spray drift for ground application, but also includes on-field.

      ii.   If ever interested in seeing impacts of intervals as a standalone parameter, then change this to False. This could be helpful for characterization.

9.   On/off field considerations

  a.   Determinations have been made on whether a species is expected to occur on different types of use site locations (see On_Off_calls_May2021.xlsx in the input folder for specifics). These on/off use site determinations were made by EFED scientists, using the ECOS provided Species Profiles, to say whether a species is likely or unlikely to be found on a specific use site.

  b.   There are currently six general on/off categories

      i.   Cultivated crops: cult_crop

      ii.   Orchards and tree plantation uses: orchards_crops

      iii.   Pasture, alfalfa, range land uses: pastures

      iv.   Residential or developed uses: residential

      v.   Forest, Managed Forest and Other Forest Tree uses: forest

      vi.   Right of way uses: row

    c.   Set any parameters to True if there is a registered use for the chemical in the category, otherwise set the parameter to False.

    d.   To help determine which UDLs are associated with the different on/off categories, see the uses-look up table (found in the inputs folder).

10. Chemical use look-up

    a.   This table identifies the specific uses registered for your chemicals and sets standard column headers for the output tables.

    b.   The easiest way to build a chemical specific table is start with the Uses_lookup_June2021_readme.xlsx file found in the input folder. Save a copy of the table as a .csv file and remove the UDLs that aren't pertinent to your chemical. When creating the .csv, be sure to associate the table with your chemical in the file name.

        i.   You can edit the information within the example based on the specifics of your chemicals.

             1.   Do not edit Column A: FullName

    c.   After generating your chemical .csv, save it to the input directory

        i.   Update the .csv file name associated with the look_up_use parameter; in single quotes

        ii.   Do not change the following text in the script: look_up_use = input_dir + os.sep +

        iii.   Final parameter:

             1.   look_up_use = input_dir + os.sep + 'Uses_lookup_June2021_[your chemical name].csv'

    d.   The UDL_LabelUse_Examples.xlsx file in the input directory is a resource to help you determine which UDLs are pertinent to your chemical.

        i.   As a general note, it is easiest to start with agricultural uses in the CONUS. If you have questions about which UDLs are most representative of your chemical's label uses, reach out to the EFED GIS team.

11. Location of the chemical independent overlap and standard species input table

    a.   The paths in this section point to the pre-processed overlap results and standard species input files download with the input directory. They are found in the sub-folders found of the input directory.

        i.   If you did not update or move files/folder within the input directory, then you will not need to make any updates to these paths. However, it is a good idea to verify them.

    b.   If you did update files/folder names to any of the standard input files, be sure to update the file names in the script accordingly.

12. Master species list

    a.   A table of species that will be included in the resulting outputs. This table is provided with the input.zip download and is in the root of the input directory.

    b.   You should not need to edit this parameter unless you change the name of the species tables. It is a good idea to verify the table name in your input directory matches the table name for this parameter.

        i.   master_list = input_dir + os.sep + 'MasterSpeciesList_Feb2022_20220216.csv'

  c. Keep in mind, the species list is updated approximately once a year, so the file name may change overtime. But this will be updated in the paths.py and input directory with each update.

  d. The col_included_output parameter identifies the columns found in the species list that will be transfer to your output table.

   i. You can delete column headers from this parameter if you don't wish to include them in your outputs.

  e. on_off_excel parameter identifies the name of the Excel spreadsheet with the on/off determination. No edit is required unless you make a change to the table/file name.

13. The paths.py script is complete. Do not edit any further.

14. Save your file.

15. Click "Run" and "Run module".

  a. You should receive a line stating RESTART: "path to paths file". This signifies that your paths to the files needed to run the overlap analysis. You are now ready to run the overlap.

Editing Scripts and Running the Overlap

You should save a copy of the scripts locally (i.e., your personal OneDrive, external hard drive, or C: drive). This will help with version control, safeguard the files, and ensure less runtime issues when running Python scripts.

There are 6 scripts within the UDL_Overlap_Scripts.zip file. The paths.py script will be the only script you will need to edit, as described in the previous section. The remaining five scripts will be run in numerical sequence once the updates to the paths.py file are complete.

You will need to run scripts 1-5 separately for the species range critical habitat. The range_run parameter found in the paths.py script allows you to move between running species range and critical habitat.

1_TabulateBySpecies.py- Extracts the results for the UDLs pertinent to your assessment from the input directory and summarizes them by species and use.

1. Right click 1_TabulateBySpecies.py and click "Edit with IDLE".
2. Click "Run" and "Run module".
3. If you receive an error, the printed message should help you find the parameter in the paths.py that should be updated.
4. If you receive a MemoryError, you are not running a 64-bit version of IDLE. See the "Software Installation Instructions" section for guidance on how to open .py files in the 64-bit version of IDLE.
5. If you are not able to solve the issue, take a screen shot of the message and send it to a member of the EFED GIS team.
6. If you do not receive any errors and do receive code stating, "End time" and "Elapsed time", the run was successful, and you can move to the next step.

2_Summarize_byPolBoundary.py- Applies the regional and state restriction based on you chemical label as defined in path.py

1. Right click 2_Summarize_byPolBoundary.py and click "Edit with IDLE".
2. Click "Run" and "Run module".
3. If you receive an error, the printed message should help you find the parameter in the paths.py that should be updated.
4. If you receive a MemoryError, you are not running a 64-bit version of IDLE See the "Software Installation Instructions" section for guidance on how to open .py files in the 64-bit version of IDLE.
5. If you are not able to solve the issue, take a screen shot of the message and send it to a member of the EFED GIS team.
6. If you do not receive any errors and do receive code stating, "End time" and "Elapsed time", the run was successful, and you can move to the next step.

3_CreateOverlapTables.py- Generates the percent overlap tables by species and use with the off-site distances set in paths.py

1. Right click 3_CreateOverlapTables.py and click "Edit with IDLE".
2. Click "Run" and "Run module".
3. If you receive an error, the printed message should help you find the parameter in the paths.py that should be updated.
4. If you receive a MemoryError, you are not running a 64-bit version of IDLE See the "Software Installation Instructions" section for guidance on how to open .py files in the 64-bit version of IDLE.
5. If you are not able to solve the issue, take a screen shot of the message and send it to a member of the EFED GIS team.
6. If you do not receive any errors and do receive code stating, "End time" and "Elapsed time", the run was successful, and you can move to the next step.

4_CreateOnOff_Tables.py- Applies the on/off site adjustment to the overlap based on chemical inputs set in paths.py.

1. Right click 4_CreateOnOff_Tables.py and click "Edit with IDLE".
2. Click "Run" and "Run module".
3. If you receive an error, the printed message should help you find the parameter in the paths.py that should be updated.
4. If you receive a MemoryError, you are not running a 64-bit version of IDLE. See the "Software Installation Instructions" section for guidance on how to open .py files in the 64-bit version of IDLE.
5. If you are not able to solve the issue, take a screen shot of the message and send it to a member of the EFED GIS team.

6. If you do not receive any errors and do receive code stating, "End time" and "Elapsed time", the run was successful, and you can move to the next step.

5_SummarizeIntervals.py- Pulls all uses for your chemical into one table, with the off-site distances set in paths.py and exports the summary overlap tables with and without the on/off site adjustments.

1. Right click 5_SummarizeIntervals.py and click "Edit with IDLE".
2. Click "Run" and "Run module".
3. If you receive an error, the printed message should help you find the parameter in the paths.py that should be updated.
4. If you receive a MemoryError, you are not running a 64-bit version of IDLE. See the "Software Installation Instructions" section for guidance on how to open .py files in the 64-bit version of IDLE.
5. If you are not able to solve the issue, take a screen shot of the message and send it to a member of the EFED GIS team.
6. If you do not receive any errors and do receive code stating, "End time" and "Elapsed time", the run was successful, and you can move to the next step.

You will need to run scripts 1-5 separately for the for the species range critical habitat. To move between species range and critical habitat update the range_run parameter in the paths.py script, see the above description.

Tool Outputs

File structure of the output directory

Output folders containing the results for range and critical habitat can be found in your output directory after you have executed the 5 scripts numerical scripts. The output directory is set using the output_dir parameter in the path.py

There will be separate output schemas for range and critical habitat. The standardized schema found within the "Range" and "Critical Habitat" folders will look as follows:

- Overlap_byCountyMerge
- Overlap_byCountyPercentOverlap
- ParentTables
- SummaryTables_1
- Tabulated_species

All of the final summarized table will be found the SummaryTable_[run number] folder. Each time you execute the tool, a new Summary Table folder is generated with a new run number. The tables found in the other folder contain intermediate data. This data is provided for reference but will not likely be needed when summarizing results.

The user defines the types of output tables that will be generated in the paths.py script using the grouping_col parameter. Three types of overlap tables can be generated, and each type of overlap will include two "scenarios", one with the on/off site adjustment and one without. The "scenario" is noted in the file name. All overlap tables will be found in the SummaryTables folder. The output options include:

1. Species range or critical habitat as a whole by use
   a. Filename without on/off adjustment ends with: [date]_SummaryIntervals.csv
   b. Filename with on/off adjustment ends with: OnOff_SummaryIntervals.csv
2. Species range or critical habitat by use and state
   a. Filename without on/off adjustment ends with: State_SummaryIntervals.csv
   b. Filename with on/off adjustment ends with: State_OnOff_SummaryIntervals.csv
3. Species range or critical habitat by use and county
   a. Filename without on/off adjustment ends with: Cnty_SummaryIntervals.csv
   b. Filename with on/off adjustment ends with: Cnty_OnOff_SummaryIntervals.csv

The file naming structure for the outputs will be the same for all chemicals and includes:
- "Range" or "CriticalHabitat", followed by the chemical identifier, "AllUses", if overlap type (i.e., "State" or "Cnty"), the date it was generated, the on/off call designation, and whether it is a summary interval file (i.e., the distances from the field that were included in the script above).

Here is an example results file:
- Range_ChemicalX_AllUses_20210624_OnOff_SummaryIntervals.csv

The data within both the range and critical habitat output folders contain intermediate data. This data is provided for reference but will not likely be needed when summarizing results. The data may be useful when incorporating specific characterization for specific uses.

Output tables

When looking at your output, the first suite of columns come directly from the Master Species List. This set of columns can be adjusted before you run the overlap as needed.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
		EntityID	Common	Scientific	ESA Listing	pop_abbr	family	Lead Agen	Group	Des_CH	CH_GIS	Source of	WoE Sum	Source of	Critical_H	Migratory	Migratory	CH_Filena	Range_Fil	L48/NL48
7		8	Florida pa	Puma (=Fe	Endangere	Wherever	Felidae	1	Mammals	FALSE	No		Terr WoE	Mammals	No Critica	No	FALSE	No	0	R_8_poly	CONUS
8		9	Sonoran p	Antilocap	Endangere	Wherever	Antilocap	1	Mammals	FALSE	No		Terr WoE	Mammals	No Critica	No	FALSE	No	0	R_9_poly	CONUS
9		11	Gray wolf	Canis lupt	Endangere	U.S.A.: All	Canidae	1	Mammals	Not Prude	No		Terr WoE	Mammals	No Critica	Yes	FALSE	No	0	R_11_poly	CONUS
10		12	Gray wolf	Canis lupt	Threatene	MN	Canidae	1	Mammals	TRUE	Yes		Terr WoE	Mammals	Terr WoE	Yes	FALSE	No	CH_12_pc	R_12_poly	CONUS
11		13	Mexican v	Canis lupt	Endangere	Wherever	Canidae	1	Mammals	FALSE	No		Terr WoE	Mammals	No Critica	No	FALSE	No	0	R_13_poly	CONUS
12		14	Red wolf	Canis rufu	Endangere	Wherever	Canidae	1	Mammals	FALSE	No		Terr WoE	Mammals	No Critica	No	FALSE	No	0	R_14_poly	CONUS

After the standard species information, you then have a suite of overlap columns, one for each use and user defined off-site distances. In all of the overlap tables, the use columns will follow same format: Region_UseName_Distance. Distance 0 represents the specific use site. Distance intervals are set by the team and might not match those in these examples.
- For instance, in the example below using range data, the output contains overlap information for corn and cotton in the conterminous US at 0, 30 and 60 m.

	A	B	C	U	V	W	X	Y	Z
1		EntityID	Common Name	CONUS_Corn_0	CONUS_Corn_30	CONUS_Corn_60	CONUS_Cotton_0	CONUS_Cotton_30	CONUS_Cotton_60
9	7	8	Florida panther	0.086754835	0.147585936	0.212270809	4.50E-06	1.95E-05	4.65E-05
10	8	9	Sonoran pronghc	0.686990038	0.887220189	1.080691246	1.523831572	1.927614752	2.312490538
11	9	11	Gray wolf	10.30881131	13.88614614	16.93040891	0	0	0
12	10	12	Gray wolf	5.92321717	8.363309695	10.61067983	0	0	0
13	11	13	Mexican wolf	0.07324354	0.121282449	0.171876397	0.03494572	0.060923806	0.08695466
14	12	14	Red wolf	0	0.01047358	0.027929548	0	0.006982387	0.020947161

Each value within the overlap columns represents the percent overlap of the species or critical habitat with each use, at each distance interval selected by the team. The percent overlap is calculated by taking the area of the use plus the area associated with off-site movement and dividing by the total area of the species location. The following equation shows how the percent overlap is calculated:

$$Percent\ Overlap_{use} = \frac{Area\ of\ use\ found\ within\ the\ species\ location}{Total\ area\ of\ the\ species\ location}$$

For example, 0.69% of the Sonoran pronghorn's range overlaps with corn at 0 m (on-field). At the next distance interval of 30 meters there is 0.89% overlap. This calculation includes the area associated with distance 0 (on-field) plus the area impacted by off-site movement up to 30 meters from the field. As you move off the field, the overlap values are inclusive of all overlap values in front of it.

If you run state or county overlaps, the output tables will contain the same information that is found above, but it is broken down by state (STUSPS) and/or county (GEOID; see below). The overlap percentages will generally be smaller in these tables if species occurs in many states and counties.

A	B	C	U	V	W	X	Y	Z	AA	AB
	EntityID	Common	GEOID	STUSPS	CONUS_Corn_0	CONUS_Corn_30	CONUS_Corn_60	CONUS_Cotton_0	CONUS_Cotton_30	CONUS_Cotton_60
1274	8	Florida pa	12086	FL	0.000797746	0.002582177	0.005044391	0	0	0
1275	8	Florida pa	12093	FL	0.001998863	0.004772978	0.008472298	0	0	0
1276	8	Florida pa	12097	FL	0.000145454	0.000427364	0.000842731	0	0	0
1277	8	Florida pa	12099	FL	0.016988083	0.027610695	0.039285972	4.50E-06	1.95E-05	4.65E-05
1278	8	Florida pa	12105	FL	6.60E-05	0.000257918	0.00059531	0	0	0
1279	8	Florida pa	12111	FL	0.034258075	0.049244296	0.06039624	0	0	0
1280	8	Florida pa	12115	FL	2.10E-05	8.85E-05	0.000209933	0	0	0
1281	9	Sonoran p	4012	AZ	0.001161283	0.002076183	0.003063356	0.056209709	0.082216206	0.104927747
1282	9	Sonoran p	4013	AZ	0.226543805	0.306077726	0.382269059	0.332999124	0.445599223	0.560810977
1283	9	Sonoran p	4019	AZ	0.022136649	0.032560272	0.043644201	0.167384075	0.19742796	0.225881857
1284	9	Sonoran p	4021	AZ	0.404668515	0.493034429	0.575084738	0.787696432	0.951522743	1.099040174
1285	9	Sonoran p	4023	AZ	0.000282519	0.000990458	0.00205483	0.001819946	0.00416715	0.0070334
1286	9	Sonoran p	4027	AZ	0.032197268	0.05248112	0.074575063	0.177722286	0.24668147	0.314796383

Interpreting the overlap

When you are ready to set the baseline overlap category, look across each of the chemicals uses plus the potential off-site transport area and set the baseline category based on the highest use overlap (including the off-site transport area) for each species. If the overlap across all uses is <1%, this species generally receives a No Effect Determination. If your overlap category is low (<5% for all use), this species would be No Jeopardy.

Overlap categories, based on USFWS Malathion Biological Opinion, include:

- Low: <5%
- Medium 5-10%
- High >10%

You should identify the number of uses with >5% for each species to support the identification of use drivers by species and uses requiring use specific mitigations before a No Jeopardy prediction can be reached.

For species with medium and high overlap, additional considerations should be made regarding the species vulnerability, and Magnitude of Effect categories prior to making a No Jeopardy or Jeopardy prediction. There also may be additional refinement/characterization possibilities for use specific overlaps prior to finalizing the use drivers, uses in need of specific mitigation, and a No Jeopardy or Jeopardy determination.

Key Considerations and Assumptions (See **Appendices A - C** for additional information):

1. **1% Overlap Limit:** Any species or critical habitat that has a greater than 1% overlap (you will use >0.44% for the purposes of rounding) with the use site and offsite distances of interest will be considered in the Jeopardy/Adverse Modification (JAM) prediction portion of the assessment. EPA has used this 1% overlap criteria as a No Effect determination because 1% is within the error bounds of spatial datasets when considering accuracy and precision.
   ii. The National Standard for Spatial Data Accuracy outlines the accepted method for calculating the accuracy of a spatial dataset (FGDC, 1998). The original GIS datasets (CDL) reports positional accuracy to 60 meters with thematic accuracies varying between 3% and >90%. The UDLs collapse the individual CDL classes to increase thematic accuracy, targeting a 90%. As a result of these accuracy assessments, and in order to prevent false precision in the analysis, the percent overlap used in decision making should be presented as whole numbers (after rounding), to account for significant digits.
   iii. For the purposes of rounding, a percent overlap that is <=0.44% is treated as a 0%, where percent overlap >=0.45% is treated as 1%. See **Appendices A - C** for additional information related to limitations/error of these remotely sensed dataset and the associated accuracy assessments.
1. **Redundancy and conservatism across UDLs:** Given the number of parent sources, temporal aggregation and other adjustments, many of the UDLs overlap with each other. Meaning a single location could be identified as a corn field, a developed layer, and within the bounds of a right of way layer. The layers are not completely independent from each other. For this reason, overlap values should not summed, and a summed result does not equate to a total overlap. See **Appendices A and B** for additional information on how the UDLs are generated.
2. **100% usage:** It is assumed every potential use site is treated.

3. **Overestimation:** Given the process for generating the UDLs, acres are often an overestimate of what is expected for a given year. The agricultural UDLs include five years of data and non-agricultural are often inclusive of area not represented in a given label use. Additionally, species locations are often an overestimation of where the species are found. It is more a representation of where a species could be found. Characterization of how these overestimations impact your determinations should be included in your assessment. Reach out to the EFED GIS team with questions and support on these types of characterizations.

Summary of Pre-Processed Data Standard Inputs Found in the Inputs Directory

Definition of Species Range and Critical Habitat

Species range is defined by the Service (i.e., US Fish and Wildlife Services [USFWS] and the National Marine and Fisheries Services [NMFS]) that is responsible for that species. A species range is organized into land areas of multiple sides or polygons. Polygons are not necessarily four-sided, and there may be many, non-contiguous, polygons the comprise a species' range.

a. The species range includes locations where the species has been identified as being present or possibly being present. Specificity of this information can range from the species historical locations, to political or man-made boundaries, to a highly analytical and refined location.

Critical habitat is a separate designation and is the location of habitat that the Services deemed as essential for the continued existence of species. Critical habitat does not need to be occupied by the species at the time of designation and may be found outside of the current species range.

a. Designated Critical Habitat includes *(1) the specific areas within the geographical area occupied by the species, at the time it is listed in accordance with the provisions of section 4 of the Act, on which are found those physical or biological features (constituent elements) (a) essential to the conservation of the species and (b) which may require special management considerations or protection; and (2) specific areas outside the geographical area occupied by the species at the time it is listed in accordance with the provisions of section 4 of the Act, upon a determination by the Secretary that such areas are essential for the conservation of the species. [ESA §3 (5)(A)] Designated critical habitats are described in 50 CFR §17 and 226."*

Spatial Inputs

The following spatial inputs are incorporated into the pre-process results found in the input directory:

- Use Data Layers (UDLs)
- Union input file(s)
- Projection files
- Snap rasters
- Political boundaries

Use Data Layers

Use Data Layers (UDLs) spatially represent areas where pesticide application could occur for agricultural and non-agricultural labeled uses. Layers are generated using a variety of land use and land cover datasets typically acquired from remote sensing technology. The spatial data used to generate these layers represent the best available landcover and land use data sets in the contiguous United States (ConUS) and the non-contiguous states and territories. A UDL represents anywhere a use site could be, not where the pesticide is known or expected to be applied (i.e., usage is assumed to be 100%).

Summary of data sources representing the spatial footprints of potential Agricultural UDLs:

- CONUS
  - Based on 5 years of USDA's Cropland Data Layer (CDL): annual raster, geo-referenced, crop-specific land cover data layer.
  - Native CDL: More than 140 crops with accuracy ranging 3% to >90% accuracy nationally.
  - UDLs: Collapses into 13 crop classes with accuracy ranged from 80% to 99% nationally.
- NL48
  - Alaska and Puerto Rico: United States Geological Survey National Land Cover Data (NLCD)
  - Hawaii, American Samoa, Guam, Commonwealth of the Marianas and the US Virgin Islands: National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP)

Data sources representing the spatial footprints of potential Non-Agricultural UDLs:

- Lot of different data sources but targeted nationally available datasets that followed the data standards outlined by the Federal Geospatial Data Committee (FDGC). 72 of the 114 non-agricultural UDLs use one of these datasets:
  - United States Geological Survey National Land Cover Data (NLCD): CONUS and NL48
  - National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP): NL48
  - United States Geological Survey LandFire Existing Vegetation Type (USGS LandFire EVT): CONUS
  - United States Geological Survey GAP Protected Areas Database (USGS GAP PAD-US): CONUS
  - United States Census Bureau's Topologically Integrated Geographic Encoding and Referencing database (TIGER): CONUS and NL48
- Total non-agricultural UDLs:
  - CONUS: 18 existing layers
  - NL48: 96 Existing layers: 96
    - Every layer may not be found in each region. Specifically, in AS, GU, CNMI, where remote sensed data is more limited.
  - Generally, the CONUS and NL48 align with each other.

Species Union Input Files

Each assessment will consider the current set of federally listed endangered and threatened species, as well as the current set of designated critical habitats. U.S. Fish and Wildlife Service and National Marine Fisheries Service provide data on listed species, as they are the experts on listed species range and biology, and thus represent the source of best available data on listed species. The Services provide spatial range and critical habitat data for each currently listed species. Those data are periodically updated in the USFWS Environmental Conservation Online System (ECOS) and on the NMFS website. Generating a current and accurate list of listed species (i.e., threatened and endangered species) is a moving target. Currently, EFED aims to update the listed species list at least once a year. The species union input files for each taxa's range and critical habitat were used to generate the pre-processed results used by this tool (found in the input directory).

Projection Files

All spatial data, such as use sites and species location (i.e., ranges and critical habitat) are projected into representative coordinate systems by region, selected to preserve area calculations. Projections flatten a 3-D globe onto a 2-D plane. Due to this flattening, there will be stretching and compressing that will affect area, location, and direction. Analysts generally choose to use a projection that minimizes distortions in the area of focus. For the overlap analysis, seven representative projected coordinate systems are used, each one specific to the different regions under U.S. jurisdictions. These regions include the contiguous United States (ConUS), Alaska (AK), Hawaii (HI), Puerto Rico (PR), United States Virgin Islands (VI), American Samoa (AS), Guam (GU), and the Commonwealth of the Northern Mariana Islands (CNMI). Each region has a unique representative projection with the exception of Guam and the Commonwealth of the Northern Mariana Islands that share the same projection. All input projection files are provided with the designated input files.

Snap Rasters

A specific setting in ESRI spatial analyst tool, this input is a raster dataset, used across analysis as a template grid. All data are aligned or snapped to the regional snap raster when ESRI tools are executed. Snap rasters are used to set the overlap extent of the analysis.

Political Boundaries

The GIS component of the analysis uses political boundaries, including state or county boundaries released by the US Census Bureau in 2021.

Tabular Inputs

The following tabular inputs will be provided:

- Master species list
- Species and critical habitat supporting tables

- o Acres
- o On/Off Calls
- Use look-up table

Most of the tabular inputs are standalone and will be used across all chemicals. The Use lookup table will need to be updated to contain only the uses of the chemical of interest.

Master Species List

The Master Species List includes the current set of federally listed (i.e., endangered and threatened) species, as well as the current set of designated critical habitats provided by The Services. This file also includes additional details on where the species and critical habitats are located and information from the weight of evidence.

Species and Critical Habitat Supporting Tables

Species and critical habitat supporting tables are also included in the inputs for the overlap analysis. These standard tables include total acres tables for the species and on/off determinations, but are not limited to these. If additional information is needed, you can work with the EFED GIS team to incorporate this information. The acres tables for both species and critical habitat include the numbers of acres across the following regions: the contiguous United States (ConUS), Alaska (AK), Hawaii (HI), Puerto Rico (PR), United States Virgin Islands (VI), American Samoa (AS), Guam (GU), and the Commonwealth of the Northern Mariana Islands (CNMI). The on/off determination are included for 6 general categories: cultivated crops, pasture, orchard/plantation, residential, right-of-way (ROW) and forest.

Use Look-Up Table

The use look-up table is the only input table that you will make edits to. This file contains the possible uses that have been pre-processed and included in the input directory. This table also aligns a given UDL with the on/off categories that were discussed previously. The use look-up table is broken out by regions.

The easiest way to build a chemical specific table is to save a copy of the provided input as a .csv file. Be sure to save the file with the associated chemical name in the file name. The name of this file will need to be updated in the paths file (see "Standard Inputs" section for more information). Then, within the file, any use that is not pertinent to the chemical of interest can be deleted. As a general note, it is easiest to start with agricultural uses in the CONUS. If you have questions about crop groupings, reach out to the EFED GIS team.

APPENDIX A: Generation of the ESA Agricultural Use Data Layers (UDLs) from the Cropland Data Layer (CDL)

Use Data Layers (UDLs) spatially represent application sites for agricultural and non-agricultural label uses in EPA's Endangered Species Biological Evaluations (BEs). They leverage several different landcover and land use datasets acquired from remote sensing technology to create a spatial footprint for a given label use. EPA uses USDA's Cropland Data Layer (CDL) for the agricultural use sites found in the conterminous United States. Updated annually, this publicly available dataset includes a robust accuracy assessment which is used by EPA to ensure the UDLs used in the BEs are of sufficient accuracy for decision making. This document provides a brief history of how this remotely sensed data is assessed for accuracy, introduces key topics related to assessing remotely sensed data, and outlines the criteria used by EPA when generating the agricultural UDLs and finally outlines the UDLs used in threatened and endangered species assessments.

Introduction to Accuracy Assessments

When selecting data sources to use to create its UDLs, EPA prefers to use publicly available national level datasets; however, it may use proprietary data if it cannot identify appropriate publicly available data. By using existing datasets, EPA leverages the expertise of other agencies and organizations, rather than becoming a 'data maker'. Generally, the selected datasets follow national standards for the creation of spatial data and, in the case of remotely sensed data, include accuracy assessments. Accuracy assessments provide a measure of correctness for the data layer. Without this measure of understanding in the spatial layers, decisions based on the dataset may lead to unexpected and possibly unacceptable results (Congalton 2019). The goal of a quantitative accuracy assessment is to identify and measure map errors so that the map can be as useful as possible to the persons making decisions. Two distinct types of quantitative accuracy assessments exist for spatial data: positional and thematic. Positional accuracy deals with the locational correctness of a map feature by measuring how far a spatial feature on a map is from its true or reference location on the ground (Bolstad 2005). The Federal Geographic Data Committee (FGDC) produced the U.S. National Cartographic Standards for Spatial Accuracy (NCSSA) (FGDC 1998) to create positional accuracy standards for medium- and small-scale maps/data. When possible, EPA leverages datasets adhering to these standards. Thematic accuracy deals with the labels or attributes of the features in the resulting GIS product and will be the focus of the discussion in this document. The thematic labels or attributes are the specific cover classes assigned in the landcover dataset. Each landcover dataset targets specific types of landscape features. In the case of the UDLs, and the underlying CDL, the primary goal of the datasets is to identify cover classes that represent agricultural crops. Other remotely sensed products may target but are not limited to non-agricultural features, non-agricultural plant cover, or water features. Each of the remotely sensed products may use the same satellite imagery, but due to the different goal of each project, the end results can differ. Thematic accuracy assessment provides measures of how different the mapped cover classes are from what occurs on the ground at specific reference locations. This is completed by comparing reference data, known/true classification of samples sites, and classified data for the same sample sites.

History of Map Making

Before the invention of aircraft, maps were created from human observations using survey equipment. Today, most map/data makers use remote sensing data rather than collecting data using field observations. To create the spatial data from remotely sensed data, decision trees algorithms use the imagery and information from known sites, referred to as training data, to generate the cover class classifications. These algorithms look for spectral signatures across multiple wavelengths to identify unique cover classes – in the CDL these are crop cover classes. Spectral signatures of various vegetation components include things such as canopy architecture, stem characteristics, leaf orientation, light angle, and shadowing of vegetation (Shah 2019). Even though advances in technology have provided access to remotely sensed information, field observations are still important and provide information at specific sample locations, used as known data for the decision tree, or as a reference site for the accuracy assessment, rather than providing a complete survey of the project area's map extent. Map/data making has moved to using remotely sensed data to make maps because it:

- is less expensive and more efficient than creating maps from human observations;
- offers a bird eye perspective, improving the understanding of spatial relationships and the context of our observations; and
- captures information in electromagnetic wavelengths that humans cannot see, such as the infrared portions of the electromagnetic spectrum, allowing for characterization of the landscape a human could not achieve.

However, no remotely sensed dataset is perfect. It is not possible to reach a complete one-to-one correlation between variation in remotely sensed data and the true variation found on the landscape. This means no resulting dataset will be error free. Several factors influence errors occurring in remotely sensed data, including but not limited to aircraft movement, topography, lens distortions, and other environmental factors (e.g., shadows, clouds, forest cover, snow morphology). These influences can reduce the strength of the relationships between the remotely sensed data and the landscape. However, errors are not limited to remotely sense datasets. The historical method of field observation also included errors due to factors such as observer bias, equipment malfunctions, inaccuracies from sampling errors, or goals of the projects.

Regardless of the collection method, no dataset will be error free. The accuracy assessment allows for an understanding of those errors and provides the user the necessary information to decide if the accuracy level meets their decision-making needs. As discussed above, remotely sensed data typically includes two types of accuracy assessment: positional and thematic. The use of remotely sensed data requires an understanding of both.

Positional accuracy is assessed by comparing the coordinates of sample/reference points on a map against the coordinates of the same points derived from a survey or some other independent source. The Federal Geographic Data Committee (FGDC) produced the U.S. National Cartographic Standards for Spatial Accuracy (NCSSA) (FGDC 1998) to create positional accuracy standards for medium- and small-scale maps/data. When possible, EPA leverages datasets adhering to these standards.

Unlike positional accuracy, there is no government or professional society standard for assessing thematic accuracy. This omission is partially due to the inherent complexity of thematic accuracy but primarily because historically, thematic accuracy was generally assumed to be at acceptable levels (Congalton 2019). The following sections explores the history of thematic accuracy and the accuracy goals set by EPA for the UDLs in absence of the government or professional society standard.

History of Thematic Accuracy

The history of assessing thematic accuracy of maps derived from remotely sensed data is relatively brief, beginning around 1975 and was divided into four parts or epochs by Congalton in 'Assessing the Accuracy of Remotely Sensed Data' (2019). Initially, no real accuracy assessment was performed on maps; rather, a "it looks good" mentality prevailed. This approach is typical of a new, emerging technology in which everything is changing so quickly that there is no time to assess how well you are doing. Despite the maturing of the technology over the last half century or so, some remote sensing analysts and map users still lean heavily on this mentality.

The second epoch is called the age of non-site-specific assessment. During this period, total acreages for each cover class were compared between reference estimates and measured without regard for location. It did not matter whether you knew where it was; only the how similar the total amounts were when compared. While total acreage is useful, it is equally if not more important to know where a specific landcover exists. Therefore, this second epoch was relatively short-lived and quickly led to the age of site-specific assessments.

In a site-specific assessment, reference locations for cover classes are compared with the classified cover class at the same location and result in a measure of overall accuracy across all cover classes in the form of a 'percent correct'. This method far exceeded the non-site-specific assessment but lacked information on individual landcover categories. Site-specific assessment techniques were the dominant method until the late 1980s.

The fourth and current age of accuracy assessment is called the 'age of the error matrix'. An error matrix compares cover class information for a number of reference sites to the remotely sensed cover class results for the same location, across each cover classes in the data layer. The error matrix is a square array of numbers set out in rows and columns, accounting for each of the cover classes. Generally, the reference data cover classes are represented as the columns and the remotely sense/classified cover classes are represented by the rows. The number in each cell represent the sample sites in the corresponding cover classes from the reference data and the classified data. The major diagonal of this matrix identifies the sites where the reference and classified cover classes match, meaning the classified data correctly identified the cover class. (**Figure A-1**).

Some key terminology when considering these matrices:
- Reference data cover classes: the class label of the accuracy assessment site derived from field or human collected data, assumed to be correct
- Classified data cover classes: the class label of the accuracy assessment site derived from the remotely sensed data.
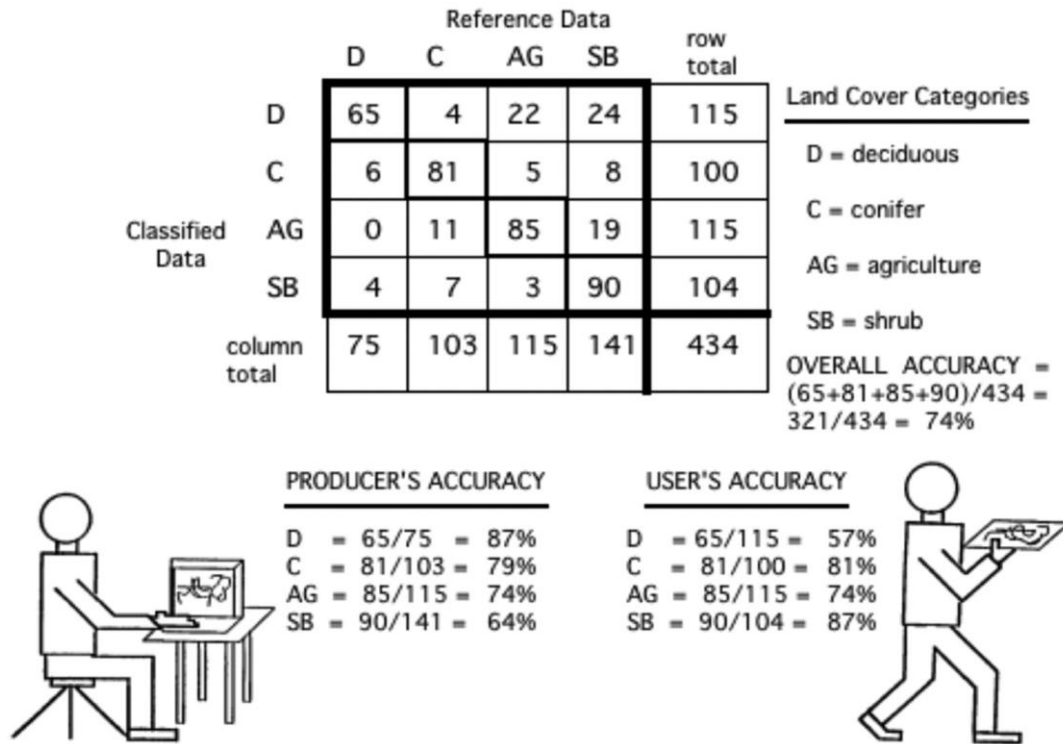
Figure A-1. Example Error Matrix and Accuracy Values (Congalton 2019). Numbers within the bolded section of the matrix are the total number of sample sites that were identified for each cover class. In this example there are a total of 434 sample sites. The number in each cell represents the total number of sample sites found with the corresponded reference and classified cover class. For example, the 65 in the top left corner indicates that 65 samples site were identified as "D" for deciduous in both the reference and classified data. However, 65 does not account for all "D" sample sites in either classified or reference data. Moving over once cell to right, there are 4 sample sites identified as "C", conifer, in the reference data but "D" in the classified data. The classified data misidentified the cover class by including it in the incorrect category – this is an error of commission. Moving down to the cell directly below 65, there 6 sites known to be "D" from the reference data but "C" in the classified data -- here the misidentified cover class results in the exclusion from a category or an error of omission. The diagonal of the error matrix represents the number of sample sites matching in the reference and classified data. The column total provides the number of sample sites found each cover classes based on the reference data, and the row total provided the number of sample sites found in each cover class based on the classified data.

With each annual release of the CDL, USDA provides error matrices for their thematic classification of cultivated land at both the national and state level.

The next sections provide additional details on the types of reported accuracy metrics provided with the error matrices, how the matrices are collapsed, and accuracy metrics are recalculated to represent the agricultural UDLs. Along with these descriptions is an example of the use of these metrics as outlined in **Figure A-1.**

Error Matrices, Overall, Producer's, and User's Accuracies, Kappa Statistic

Error matrices are effective representations of map accuracy, because the individual accuracies of each map cover class are plainly described on the major diagonal (i.e. classified data that matches the reference data), along with both the errors of inclusion (also referred to as "commission errors") and the errors of exclusion (also referred to as "omission errors") when the classified and reference data cover classes do not match. An omission error occurs when a sample site is left out, or omitted, from the correct classes in the classified dataset. This is considered a false positive of the classified data or Type 1 error. A commission error occurs when a sample site is included in an incorrect class in the classified dataset. This is considered a false negative/false match of the classified data or Type 2 error.

In addition to clearly showing errors of omission and commission, the error matrix can be used to compute overall accuracy, producer's accuracy, and user's accuracy, which were introduced to the remote sensing community by Story and Congalton (1986). Overall accuracy is simply the sum of the major diagonal divided by the total number of sample units, providing a 'percent correct' across all cover classes. In the example error matrix found in **Figure A1**, the overall accuracy is the sum of the values on the major diagonal, where the classified and reference data match, divided by the total number of sample sites or 321/435; resulting in an overall accuracy of 74%. This value is the most commonly reported accuracy assessment statistic. In addition to the overall accuracy, the reporting of producer's and user's accuracies allow for additional considerations, specifically of individual cover classes.

Computed to determine individual cover class accuracies, producer's and user's accuracies provide important information related to error within the individual cover class from different perspectives. The producer of the map may want to know how well a class matched the reference data, referred to the producer's accuracy. This value is computed by dividing the value from the major diagonal (the agreement between the reference and classified data) for the class of interest, by the total number of reference data points for the class. Looking at **Figure A-1**, the map producer identified 65 sites as deciduous, while the reference data indicate there were a total of 75 deciduous sites. So, 65 of 75 samples were correctly identified, resulting in a producer's accuracy of 87%, which is quite good. However, this is only half of the story. If you now view the map from the user's perspective, a user wants to know how many classified data points matched the reference data. In **Figure A-1,** you see once again that 65 sites were classified as deciduous on the map that were actually deciduous, but the map shows a total of 115 site classified as deciduous, resulting in a user accuracy of 57%. In evaluating the accuracy of an individual map class, it is important to consider both the producer's and the user's accuracies.

The kappa statistic or coefficient is used as another measure of agreement for the resulting remotely sensed data (Cohen, 1960). This measure of agreement is based on the difference between the actual agreement in the error matrix (i.e., the agreement between the remotely sensed classification and the reference data as indicated by the major diagonal) and the chance agreement, which is indicated by the row and column totals (i.e., marginals). The kappa reflects agreement between the classified cover classes and the reference cover classes, and ranges from 0 to 1. If the kappa equals 0 than there is no agreement between the classified and references label. The closer to 1 the kappa, the closer the

agreement is, and if it reaches 1 then the classified and reference data match perfectly. Ultimately, a Kappa of 0.85 means there is an 85% or better agreement than chance alone.

$$\hat{K} = \frac{observed\ accuracy - chance\ agreement}{1 - chance\ agreement}$$

The power of kappa is in its ability to test whether one error matrix is statistically significantly different from another and not in simply reporting this value as another measure of accuracy.

Use of Accuracy Values in Understanding Thematic Errors

In the past, an overall accuracy level of 85% was often adopted as representing the cutoff between acceptable and unacceptable data. This standard was first proposed in Anderson et al. (1976) despite the lack of any research being performed to establish this standard. Accuracy depends on many factors, including the amount of effort, level of landscape or classification detail, and variability of the classes. In some instances, an overall accuracy of 85% is more than sufficient; in others it would not be accurate enough; and in others, such an accuracy would be way too expensive to ever achieve (Congalton 2019).

In the example described above and presented in **Figure A-1**, the error matrix has an overall map accuracy of 74%. This value tells about how accurate the map is, in general or across all classes, but provides no information within individual classes. For additional information on the deciduous cover class, the producer's and user's accuracies can be considered. The producer's accuracy for this class of 87% is quite good and even higher the overall accuracy of the dataset. However, if we stopped there, one might conclude that although the dataset appears to be average overall (i.e., 74%), it is more than adequate for the deciduous class. Making such a conclusion could be a serious mistake because the user's accuracy of 57% tells a different story. In other words, although 87% of the deciduous areas have been correctly identified as deciduous, only 57% of the areas called deciduous on the map are actually deciduous based on the reference data. This lower user accuracy tells us that there are errors of commission in the map related to the deciduous classes, meaning there are sample sites that were classified as deciduous that based on the reference belong to a different class. The result of this is more area in the map classified as deciduous than actually occurs on the ground.

A more careful look at the error matrix reveals significant confusion in discriminating deciduous from barren and shrub. Therefore, although the producer of this map can claim that 87% of the time an area that was deciduous on the ground was identified as such on the map, a user of this map will find that only 57% of the time that the map says an area is deciduous will it actually be deciduous on the ground and may often be barren/scrub.

The intended use of the data/map can drive the need to address some of the error. For example, the lower user accuracy in the example above often resulted from the confusion between discriminating deciduous from barren/shrub. Collapsing these two classes together into a deciduous/barren/shrub class increase the user's accuracy to 83% but lowers the producer's accuracy to 85% (**Figure A-2**). The higher user's accuracy means when the map identifies this grouped cover class it matches what is found on the ground more often than the two individual classes. Under certain situations it may be worth the

slightly lower producer accuracy and sacrificing one of the cover classes, meaning the map will no longer distinguish between deciduous and shrub/barren.

	D/SB	C	AG	Row total
D/SB	183 (65+4 +114)	11	25	219
C	14 (6+8)	81	5	100
AG	19 (0+19)	11	85	115
Column total	216	103	115	434

$$\text{Producer's accuracy} = \frac{183}{216} = 85\% \quad \text{User's accuracy} = \frac{183}{219} = 83\%$$

Figure A-2. Example collapsing cover class to address error of commission, building off the error matrix in Figure A-1 here the deciduous and barren/shrub are combined, and accuracy metric recalculated.

For the purposes of the UDL, EPA's targets at least 85% in both the producer's and user's accuracy and at least 90% for an overall accuracy when combining individual crops from the CDL into the UDL cover classes.

EPA's Accuracy Value Goals for ESA Use Data Layers

The native CDL landcover dataset includes over 100 cultivated cover classes in its thematic classification. The error matrices released with the CDL data provide overall, producer and user measures of accuracy at both the state and national level as well as the associated Kappas. In recent years, the overall accuracy of the CDL dataset has been in the low to mid-80% with Kappa just over 0.80. The producer's and user's accuracy for the individual cultivated classes range from less than 5% to 98%, and less than 15%-97%, respectively (Boryan 2011). When considering the individual cultivated classes of the CDL, the user's accuracy is slightly better than producer's accuracy, resulting in a lower commission error, or false negative/Type 2 error. However, when considering these BEs, reducing the false positive/Type 1 error is equally or more important. Improving all accuracy metrics as well as leveling out the producer and user accuracies is an overall goal when grouping crops into the UDLs cover classes.

To improve the overall, user and producer accuracies for the UDLs, the 100+ thematic cultivated classes found in CDL are reclassified into 13 crop groupings. Consolidating CDL into aggregated categories is a documented way to significantly improve the accuracy of assessments by eliminating misclassification errors within the combined classes (Johnson 2013a, Johnson 2013b, Wright 2013 and Lark 2017). Each of the 100+ thematic cultivated classes from the original CDL, are found in at least one state but not every state will include all 100+ classes. For this reason, while the focus is on the accuracy at the national level, there are instances when the state accuracy for a UDL would be higher than observed at the national level.

When deciding how to group crops from the CDL, EPA refers to the grouping used by the U.S. Geological Survey (Baker and Capel, 2011) and the Generic Endangered Species Task Force (Amos et al 2010). This information considers environmental factors that influence the location of crops and the error matrices

provided by USDA with the original CDL data. By considering these agronomic factors in addition to the error matrices it is possible to improve the accuracy for these UDLs while retaining agronomic similarities. There is an infinite number of ways to group the crop cover classes found in the CDL, and each structured grouping can be reviewed in terms of recalculated accuracy compared to the native dataset.

When collapsing the available error matrices provided with the CDL into the 13 UDL groups, the sample site values for each of the CDL crops found in a UDL are summed across both rows and columns in the error matrix. Currently the 13 UDL groups bring the overall accuracy to 90%, increased from 80% for the CDL, with a Kappa of 0.88 (**Table A-1**). As described above, it is important to consider the producer's and user's accuracy of the individual thematic classes in addition to the overall accuracy.

When considering the user's and producer's accuracy, EPA targets at least 85% for each UDL, while retaining at least a 90% overall accuracy. Following the thematic grouping into the 13 UDLs and the recalculation of the user and producer accuracies, by year of the CDL, to help address errors of commission, additional steps to lower the omission errors, are implemented. These include the temporal aggregation of multiple CDL years into the UDL and expanding the crop area found in the UDL layer to meet or exceed the area for the same suite of crops as reported in the Census of Agriculture. The goal of each of these steps is to improve the accuracy of the UDLs by minimizing the rate of omission error. However, these steps are not directly related to the existing error matrices provided with the CDL, and therefore new accuracy values are not calculated following the temporal aggregation, and area expansion. By reducing the omission errors, these steps result in a more protective landcover classification for each UDL.

If an individual crop class in the CDL has both the producer and user accuracies that are over 85%, the corresponding UDLs is that same as the CDL crop cover class, for example cotton from the CDL is found in the cotton UDL. These UDLs include corn, cotton, grapes/other vineyards, rice, soybeans and wheat. Five of these UDLs have user and producer accuracies in the low to mid 90%, with Kappas ranging from ~0.89 to 0.97. The user's and producer's accuracy for the remaining cotton UDL falling above 85% with Kappas of ~ 0.85. Due to the geographically limited occurrence of cotton, this crop is only grown in the south, lower national accuracy is expected compared to other crops with a broader geographic range. This is due to the fact that cotton growing states may classify cotton well, however, there is a lower accuracy in identifying cotton in states where cotton doesn't grow, and this brings down the national accuracy.

When an individual crop cover class in the CDL is below 85%, grouping multiple crops together and ultimately reducing the number of total thematic crop groups, improves the accuracy of the resulting UDL. When deciding which crops to group, error of omission and commission of the remotely sensed data are considered, in addition to environmental and agronomic practices. EPA targets an accuracy of at least 85%; however, it is not always possible to reach the target without compromising the environmental/agronomic practices. For this reason, some of the UDLs that contain multiple crop classes have slightly lower than 85% accuracy.

The UDLs containing a number of crops include alfalfa/other agricultural grasses, citrus, other crops, other grains, other orchards, other row crops, and vegetables and ground fruit. Two of these UDLs, other crops and other grains, did not meet an 85% accuracy for user's and producer's accuracy. Two additional UDLs, other row crops and vegetables and ground fruit, did not reach 85% for just the producer's accuracy. See **Table A-1** for a complete list of accuracy values across all 13 UDLs. As mentioned above, the focus of the discussion is on the national accuracies. But due to the variety and regional nature of some crops found in the UDLs, state-based accuracy assessments often reach 85% even though the national level assessment for the same UDL does not.

Additional challenges when identifying some crops include higher frequency of change in agricultural practices (e.g., crop rotation), and/or lower total area on the landscape for minor crops. These two challenges are related to errors of omission, rather than errors of commission addressed by grouping crops into the UDL categories a common practice implemented to increase accuracy of remotely sensed data (Johnson 2013a, Johnson 2013b, Wright 2013 and Lark 2017). Two additional steps address some of the uncertainty related to these errors of omission, specifically, the known downward estimates of acres for remotely senses data and changes in crop patterns over time. These steps are implemented on all UDLs but have the most impact in addressing uncertainty around error of omission for the UDLs containing multiple crops with lower accuracy values. First, a temporal aggregation of multiple years of the CDL into the UDLs is performed to account for changing agricultural practices, for example crop rotation, from year to year. Second, the total area of the temporally aggregated UDL is compared to the reported area found in the Census of Agriculture, accounting for some of the error/difficulty in identifying minor crops. If the area of the UDL is less than the reported area in the Census of Agriculture, the UDL is grown out to meet or exceed the Census of Agriculture. Referred to as region growing, expanding the UDL area to meet or exceed the area reported in the Census of Agriculture is a conservative measure take to minimize the error of omission. However, the Census of Agriculture generated once every 5 years, represents a single year in time. The CDL generated every year may capture agricultural practices, such as rotations, not captured in the Census Agriculture. For this reason, there is uncertainty around the crop area found in the Census of Agriculture being representative across all years of the CDL.

At the end of the whole process, the resulting UDLs provide a more protective land cover estimates for the purposes of assessing impacts to threatened and endangered species, making them the best available agricultural spatial data to use in threatened and endangered species chemical assessments.

Figure A-3 provide a summary of the classes found in each of the ESA agricultural UDLs with a complete crosswalk of the original CDL crops to the UDL class provided in **Table A-2.**

Table A-1. Collapsed national error matrix from the 2018 CDL, example of the 13 national UDLs with associated measures of accuracy.

	1	2	3	4	5	6	7	8	9	10	11	12	13	User's Acc.	Commission	Kappa
Alfalfa	2157632	325	49580	6026	440	38838	45476	4745	4226	131	27170	13039	29148	89%	11%	0.87
Citrus	147	244865	37	25	12	185	112	103	1	0	0	164	3	99%	1%	1.00
Corn	39172	26	4222089	6598	241	18927	32759	1636	6212	1454	124498	20895	13154	94%	6%	0.92
Cotton	5368	12	9800	974234	51	9753	17664	1405	43844	509	36809	5983	15474	87%	13%	0.86
Grapes	426	30	546	35	93320	1372	47	3206	607	0	56	288	92	93%	7%	0.93
Other Crops	26196	385	12842	7554	581	729904	37343	6695	4335	2888	11038	9363	32155	82%	18%	0.82
Other Grains	16615	23	14503	7531	20	18118	597678	312	3603	210	8702	7707	34988	84%	16%	0.83
Other Orchards	2870	234	1305	1717	1862	3680	521	353321	950	26	524	1424	412	96%	4%	0.96
Other Row Crops	2528	0	3208	13781	208	2860	4999	466	315797	165	3933	2981	782	90%	10%	0.89
Rice	150	0	1061	154	1	3158	340	5	36	275819	2509	190	106	97%	3%	0.97
Soybeans	28675	0	139339	54449	101	29702	25116	427	10953	15386	4754850	16137	27339	93%	7%	0.90
Vegetables and ground fruit	5221	83	6822	1587	289	6397	7439	1209	3009	106	3263	361780	5496	90%	10%	0.90
Wheat	22383	0	11027	18183	99	44103	81911	525	1618	10	22050	13228	1833412	89%	11%	0.88
Producer's Accuracy	85%	99%	93%	89%	95%	78%	68%	93%	79%	93%	94%	79%	90%			
Omission	15%	1%	7%	11%	5%	22%	32%	7%	21%	7%	6%	21%	10%			
Kappa	0.82	0.99	0.91	0.88	0.95	0.77	0.67	0.93	0.79	0.93	0.91	0.79	0.89			

Overall Accuracy	90%
Overall Kappa	0.88

Figure A-3. Summary of ESA Use Data Layer Classes

Reclass Value	UDL General Classes
10	Corn
14	Corn/soybeans
15	Corn/wheat
18	Corn/grains
20	Cotton
25	Cotton/wheat
26	Cotton/vegetables
30	Rice
40	Soybeans
42	Soybeans/cotton
45	Soybeans/wheat
48	Soybeans/grains
50	Wheat
56	Wheat/vegetables
58	Wheat/grains
60	Vegetables and ground fruit
61	(ground fruit)
68	Vegetables/grains
70	Other Orchards
71	Vineyards
72	Citrus
80	Other grains
90	Other row crops
100	Other crops
110	Alfalfa/agricultural grasses

*These classes are not mutually exclusive to one another and are further reclassified into 13 national agricultural UDL classes.. The complete crosswalk for all 13 UDL classes can be found in **Table 2.***
Corn: 10, 14, 15, 18
Cotton: 20, 25, 26, 42
Rice: 30
Soybeans: 40, 42, 45, 48, 14
Wheat: 50, 56, 58, 15, 25, 45
Vegetables & Ground Fruit: 60, 61, 68, 26, 56
Other Orchards: 70
Vineyards: 71
Citrus: 72
Other Grains: 80, 18, 48, 58
Other Row Crops: 90
Other Crops: 100
Alfalfa/agricultural grasses: 110

Table A-2. Cross-walk between CDL class and UDL agricultural classes.

CDL Value	CDL Class Name	Reclass Category for UDLs	Double Crop (Y)	Reclass Code
1	Corn	Corn		10
2	Cotton	Cotton		20
3	Rice	Rice		30
4	Sorghum	Other grains		80
5	Soybeans	Soybeans		40
6	Sunflower	Other row crops		90
10	Peanuts	Other row crops		90
11	Tobacco	Other row crops		90
12	Sweet Corn	Vegetables and ground fruit		60
13	Popcorn Corn	Vegetables and ground fruit		60
14	Mint	Vegetables and ground fruit		60
21	Barley	Other grains		80
22	Durum Wheat	Wheat		50
23	Spring Wheat	Wheat		50
24	Winter Wheat	Wheat		50
25	Other Small Grains	Other grains		80
26	Double Crop Winter Wheat/Soybeans	Soybeans/Wheat	Y	45
27	Rye	Other grains		80
28	Oats	Other grains		80
29	Millet	Other grains		80
30	Speltz	Other grains		80
31	Canola	Other grains		80
32	Flaxseed	Other grains		80
33	Safflower	Other grains		80
34	Rape Seed	Other grains		80
35	Mustard	Vegetables and ground fruit		60
36	Alfalfa	Alfalfa/agricultural grasses		110
38	Camelina	Other grains		80
39	Buckwheat	Other grains		80
41	Sugarbeets	Other row crops		90
42	Dry Beans	Vegetables and ground fruit		60
43	Potatoes	Vegetables and ground fruit		60
44	Other Crops	Other crops		100
45	Sugarcane	Other grains		80
46	Sweet Potatoes	Vegetables and ground fruit		60
47	Misc Vegs & Fruits	Vegetables and ground fruit		60
48	Watermelons	Vegetables and ground fruit		60
49	Onions	Vegetables and ground fruit		60

CDL Value	CDL Class Name	Reclass Category for UDLs	Double Crop (Y)	Reclass Code
50	Cucumbers	Vegetables and ground fruit		60
51	Chick Peas	Vegetables and ground fruit		60
52	Lentils	Vegetables and ground fruit		60
53	Peas	Vegetables and ground fruit		60
54	Tomatoes	Vegetables and ground fruit		60
55	Caneberries	Vegetables and ground fruit		61
56	Hops	Other row crops		90
57	Herbs	Vegetables and ground fruit		60
58	Clover/Wildflowers	Other crops		100
59	Sod/Grass Seed	Other crops		100
60	Switchgrass	Alfalfa/agricultural grasses		110
61	Fallow/Idle Cropland	Other crops		100
66	Cherries	Other orchards		70
67	Peaches	Other orchards		70
68	Apples	Other orchards		70
69	Vineyards	Vineyards		71
70	Christmas Trees	Other trees		75
71	Other Tree Crops	Other orchards		70
72	Citrus	Citrus		72
74	Pecans	Other orchards		70
75	Almonds	Other orchards		70
76	Walnuts	Other orchards		70
77	Pears	Other orchards		70
92	Aquaculture	Other crops		100
141	Deciduous Forest	Forest		140
142	Evergreen Forest	Forest		140
143	Mixed Forest	Forest		140
152	Shrubland	Shrubland		160
204	Pistachios	Other orchards		70
205	Triticale	Other grains		80
206	Carrots	Vegetables and ground fruit		60
207	Asparagus	Vegetables and ground fruit		60
208	Garlic	Vegetables and ground fruit		60
209	Cantaloupes	Vegetables and ground fruit		60
210	Prunes	Other orchards		70
211	Olives	Other orchards		70
212	Oranges	Citrus		72
213	Honeydew Melons	Vegetables and ground fruit		60
214	Broccoli	Vegetables and ground fruit		60
216	Peppers	Vegetables and ground fruit		60

CDL Value	CDL Class Name	Reclass Category for UDLs	Double Crop (Y)	Reclass Code
217	Pomegranates	Other orchards		70
218	Nectarines	Other orchards		70
219	Greens	Vegetables and ground fruit		60
220	Plums	Other orchards		70
221	Strawberries	Vegetables and ground fruit		61
222	Squash	Vegetables and ground fruit		60
223	Apricots	Other orchards		70
224	Vetch	Alfalfa/agricultural grasses		110
225	Double Crop Winter Wheat/Corn	Corn/Wheat	Y	15
226	Double Crop Oats/Corn	Corn/Grains	Y	18
227	Lettuce	Vegetables and ground fruit		60
229	Pumpkins	Vegetables and ground fruit		60
230	Double Crop Lettuce/Durum Wheat	Wheat/Vegetables	Y	56
231	Double Crop Lettuce/Cantaloupe	Vegetables and ground fruit		60
232	Double Crop Lettuce/Cotton	Cotton/Vegetables	Y	26
233	Double Crop Lettuce/Barley	Vegetables/Grains	Y	68
234	Double Crop Durum Wheat/Sorghum	Wheat/Grains	Y	58
235	Double Crop Barley/Sorghum	Other grains		80
236	Double Crop Winter Wheat/Sorghum	Wheat/Grains	Y	58
237	Double Crop Barley/Corn	Corn/Grains	Y	18
238	Double Crop Winter Wheat/Cotton	Cotton/Wheat	Y	25
239	Double Crop Soybeans/Cotton	Soybeans/Cotton	Y	42
240	Double Crop Soybeans/Oats	Soybeans/Grains	Y	48
241	Double Crop Corn/Soybeans	Corn/Soybeans	Y	14
242	Blueberries	Vegetables and ground fruit		61
243	Cabbage	Vegetables and ground fruit		60
244	Cauliflower	Vegetables and ground fruit		60
245	Celery	Vegetables and ground fruit		60
246	Radishes	Vegetables and ground fruit		60
247	Turnips	Vegetables and ground fruit		60
248	Eggplants	Vegetables and ground fruit		60
249	Gourds	Vegetables and ground fruit		60
250	Cranberries	Vegetables and ground fruit		61
254	Double Crop Barley/Soybeans	Soybeans/Grains	Y	48

APPENDIX B. Data Used to Generate the Use Data Layers (UDLs)

Agriculture Uses

Use data layers (UDLs) represent the application sites for agricultural and non-agricultural label uses. The best available data to spatially characterize specific agricultural crops in the continuous United States (ConUS) is the Cropland Data Layer (CDL), produced by the U.S. Department of Agriculture. Several methods have been employed to minimize data errors within the CDL. The CDL is a landcover dataset that has over 100 cultivated classes that were grouped into 13 general classes (see **APPENDIX A**). Lumping classes reduces the likelihood of errors of omission and commission between similar crop categories. In selecting how to group crops from the CDL, EPA referred to the grouping used by the U.S. Geological Survey (Baker and Capel, 2011) and the Generic Endangered Species Task Force (Amos et al., 2010). This information considers environmental factors that influence the location of crops and the error matrices provided by USDA with the original CDL data. By considering these agronomic factors in addition to the error matrices it is possible to improve the accuracy and year-to-year matches for these UDLs while retaining agronomic similarities. This categorical aggregation into the UDL crop groups does not account for changes in agricultural practices but the temporal aggregation does. The UDLs used in this assessment include 5 years of the CDL, 2013-2017, aggregated to account for changes year to year such as crop rotations. Anywhere a class occurs within those 5 years would be represented in the footprint layer. These temporally aggregated and categorially grouped layers generated from the CDL are referred to as Use Data Layers or UDLs.

The agricultural classes were further refined by comparing county level National Agricultural Statistics Service (NASS) 2012 Census of Agriculture (CoA) acreage reports to county level UDL acreages. The UDL acreages represent the temporally aggregated and categorically grouped processing steps previously described, summarized at the county level. If a county's UDL acreage for a given class was lower than the NASS acreage, the UDL extent was expanded within cultivated areas until the UDL acreage matched or exceeded the NASS CoA. Using the temporally and categorically aggregated UDL as an input, a script was developed that compares each UDL in each county to the corresponding NASS CoA acreage report. If the UDL acreage was less than NASS, the raster was expanded in 1-pixel iterations until the NASS acreage value was reached, exceeded, or the area within the cultivated mask was built out. Region growing was restricted using the UDL Cultivated Layer from the last year of the CDL as a mask (2017). This avoids buffering into any non-agricultural landcover types. This method reduces uncertainty related to landcover mapping by ensure the acres mapped on the ground in the UDL corresponds to the reported acreage from the CoA in this case, 2012. This helps addresses the uncertainty in acreage estimates in the landcover data given the known downward area bias in area estimates related to remotely sensed data.

Every assessment begins with cross-walking registered uses into a UDL.

In addition to the potential use site each UDL is buffered in all directions using the ESRI ArcGIS Euclidean distance tool. This buffered area represents the potential exposure area associated with off-site movement.

The CDL is not available for areas outside of the contiguous United States (ConUS). The CoA is often unavailable outside of ConUS as well. The Agricultural UDL Data Sources section describes how agriculture was spatially modeled by regions outside of ConUS, referred to as the Non-lower 48 (NL48).

Non-Agricultural Uses

Non-agricultural label uses include a wide range of landcover and land use categories. Each label use should be carefully considered and cross-walked with the best UDL. Where available, the 2016 National Land Cover Dataset (NLCD) was used to represent many non-agricultural label uses. Details on the data sources for each non-agricultural UDL are provided in the "Non-Agricultural UDL Data Source" section.

UDL Data Sources

Agricultural UDL Data Sources for CONUS and the NL48
ConUS

- Corn, Cotton, Soybeans, Wheat, Citrus, Grapes, Other Grains, Other Orchards, Other Row Crops, Other Crops, Alfalfa/other agricultural grasses (non-grazing areas), Rice, and Vegetables and Ground Fruit. These agriculture UDLs were generated from the Cropland Data Layer (CDL) 2013-2017. See **APPENDIX A** for details on the specific crops found in each UDL. National UDLS were used.
- Cultivated/Fallow
    - Cultivated/Fallow is spatial represented using all cultivated land as identified in USDA's Cultivated layer from Cropland Data Layer (2017). It is based on the most recent five years of CDL data. The general processing logic is as follows: if a pixel is identified as cultivated in at least two out of the five years of CDL data then it is assigned to the 'Cultivated' category. The exception is that all pixels identified as cultivated in the most recent year are assigned to the 'Cultivated' category regardless of whether or not they were cultivated in the previous four years of CDL data.

Alaska (AK)

- National Land Cover Dataset (NLCD 2016) Cultivated Class (82); inclusive of all agricultural crops, Bermuda grass pasture and fallow

Hawaii (HI)

- National Oceanic & Atmospheric Administration (NOAA) Coastal Change Analysis Program (CCAP 2011), Cultivated Class (6), inclusive of all agricultural crops, Bermuda grass pasture and fallow

Puerto Rico (PR)

- NLCD Cultivated Class (2001) (82), inclusive of all agricultural crops, Bermuda grass pasture and fallow

Guam (GU)

- CCAP 2011 Cultivated Class (6), inclusive of all agricultural crops, Bermuda grass pasture and fallow

Marianas (CNMI)

- CCAP 2004 Cultivated Class (6); inclusive of all agricultural crops, Bermuda grass pasture and fallow

American Samoa (AS)

- CCAP 2010 Cultivated Class (6) inclusive of all agricultural crops, Bermuda grass pasture and fallow

Virgin Islands (VI)

- CCAP 2012 Cultivated Class (6) inclusive of all agricultural crops, Bermuda grass pasture and fallow

Non-Agricultural UDL Data Sources ConUS and NL48 – National Landcover Dataset (NLCD)

Non-agricultural label uses include a wide range of landcover and land use categories. Each label use was carefully considered and cross-walked with the best available UDL. It is possible for a label use to crosswalked to multiple UDLs, this is discussed in more detail in the individual UDL sections below. Where available, the 2016 National Land Cover Dataset (NLCD) for ConUs and AK and 2001 NLCD in PR was used to represent many non-agricultural label uses (see below). Where NLCD wasn't available, the NOAA C-CAP and other dataset outlined below were used. Below is a sample of label classes that were represented using NLCD Developed or Open Space Developed land use categories. These different Developed classes are discussed in more detail below. Forest landcover categories from the NLCD or CCAP for forest tree label uses.

- Grain/cereal/flour bins
- Grain/cereal/flour elevators
- Household/domestic dwellings (perimeter outdoor only)
- Non-agricultural outdoor building structures
- Ornamental and/or shade trees
- Ornamental herbaceous plants
- Ornamental non-flowering plants
- Ornamental woody shrubs and vines
- Refuse/solid waste containers (outdoors)
- Refuse/solid waste sites (outdoors)
- Commercial/Institution-Al/ Industrial Premises/ Equip. (Indoor and Outdoor) – Broadcast, Crack and Crevice/Void
- Domestic Dwellings Outdoor Premises;
- Food Processing Plant Premises (Nonfood Contact) – Crack and Crevice
- Nonagricultural Outdoor Buildings/Structures
- Poultry Litter (Poultry houses)

- Recreational Areas
- Sewer Manhole Covers and Walls
- Utilities – Broadcast
- Wood Protection Treatment to Buildings/Products Outdoor

Developed

Developed land cover is used to spatially represent certain non-agricultural label uses and includes areas with a mixture of constructed materials and vegetation, where impervious surfaces account for 20% to 100 % percent of total cover. These areas most commonly include single-family housing units, apartment complexes, row houses and commercial/industrial.

- ConUS
  - NLCD 2016 class 22-24
- Alaska
  - NLCD 2016 class 22-24
- Hawaii
  - CCAP 2011 class 2-4
- Puerto Rico
  - NLCD 2001 class 22-24
- Guam
  - CCAP 2011 class 2
- Marianas
  - CCAP 2004 class 2
- American Samoa
  - CCAP 2010 class 2
- Virgin Islands
  - CCAP 2012 class 2

Open Space Developed

Open Space Developed (OSD) is used to spatially represent certain non-agricultural label uses and includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn/turf grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.

- ConUS
  - NLCD 2016 class 21
- Alaska
  - NLCD 2016 class 21
- Hawaii
  - CCAP 2011 class 5
- Puerto Rico
  - NLCD 2016 class 21
- Guam
  - CCAP 2011 class 5
- Marianas

- o   CCAP 2004 class 5
- American Samoa
  - o   CCAP 2010 class 5
- Virgin Islands
  - o   CCAP 2012 class 5

Forest Trees
- CONUS
  - o   NLCD indicates 2016 "Forest" (41-43)
- Alaska
  - o   NLCD indicates 2019 "Forest" (41-43)
- Hawaii
  - o   CCAP 2011 indicates "Forest" (9-11)
- Puerto Rico
  - o   NLCD indicates 2001 "Forest" (41-43)
- Guam
  - o   CCAP 2011 indicates "Forest" (9-11)
- Marianas
  - o   CCAP 2004 indicates "Forest" (9-11)
- American Samoa
  - o   CCAP 2010 indicates "Forest" (9-11)
- Virgin Islands
  - o   CCAP 2012 indicates "Forest" (9-11)"

Non-Agricultural UDL Data Sources ConUS and NL48 – Other Datasets

When the NLCD was inadequate to represent a label use, other data sources were used in modeling as appropriate. The following list describes each label use and how it was spatially modeled by region.

Christmas Trees
Cropland Data Layer (CDL) class 70, Christmas Trees, are used for ConUS. These are not characterized anywhere else.

ConUS

- Cropland Data Layer (CDL) class 70, Christmas Trees

All Non-Lower 48 Regions:

- No Christmas Tree land cover data is available

Conservation Reserve Program
Cultivated area that could be or have been placed into the conservation reserve program

ConUS

- Conservation Reserve Program is spatial represented using all cultivated land as identified in USDA's Cropland Data Layer (2017).

Alaska (AK)

- NLCD 2016 Cultivated Class (82)

Hawaii (HI)

- CCAP 2011 Cultivated Class (6)

Puerto Rico (PR)

- NLCD 2001 Cultivated Class (82)

Guam (GU)

- CCAP 2011 Cultivated Class (6)

Marianas (CNMI)

- CCAP 2004 Cultivated Class (6)

American Samoa (AS)

- CCAP 2010 Cultivated Class (6)

Virgin Islands (VI)

- CCAP 2012 Cultivated Class (6)

Field Nurseries

Non-agricultural Nurseries represent a land use that is not exclusive to any nationwide land cover class. Nurseries are mapped by using geocoded Dun and Bradstreet (D&B) business database addresses. Label uses that are covered by this UDL found on ornamentals, shrubs/vines, and non-food trees, grown in a non-agricultural setting (e.g., Retail Nurseries, Garden supply stores, retail greenhouse, retail shade house or retail horticultural location. This information is combined with the UDLs for Other Orchard and Citrus in ConUS to represented by agricultural nursery uses such as trees grown for food, tree plantations or transplanted trees, shrubs, and ornamentals.

ConUS

- Using the Dun and Bradstreet business database, select all records with any SIC Codes starting with "018" (Horticultural Specialties) or "526" (Retail Nurseries, Lawn nd Garden Supply Stores)
- Selected points are then buffered by their facility size attribute. Where facility size is absent, substitute the Census of Agriculture's average acreage by county, calculated using Nursery Totals. If a county's nursery acreages are undisclosed, then an average of all county averages is used. A circular buffer is applied, where radius is solved for using the areas previously described. In an effort to map production facilities only and not business offices, use the 'Location Type' attribute to categorize locations.
- Use Data Layer for Other Orchard and Citrus (CDL 2013-2017)

Alaska

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Hawaii

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Puerto Rico

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Guam

- No Dun and Bradstreet business data were available for Guam.

Marianas

- No Dun and Bradstreet business data were available for Marianas.

American Samoa

- No Dun and Bradstreet business data were available for American Samoa.

Virgin Islands

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Forestry (Managed Forest + Xmas trees)

Forested areas managed for timber extraction, forest tree plantations (including Christmas Trees). There is also a stand-alone Christmas Tree UDL, see below.

ConUS

- Cropland Data Layer (CDL) class 70, Christmas Trees
- Include all the following LandFire Existing Vegetation Type (EVT) classes; "Recently Logged-Herb and Grass Cover", "Recently Logged-Shrub Cover", "Recently Logged-Tree Cover", "Managed Tree Plantation-Northern and Central Hardwood and Conifer Plantation Group", or "Managed Tree Plantation-Southeast Conifer and Hardwood Plantation Group"
- Include any of the following United States Geologic Survey (USGS) National Gap Analysis Program (GAP) Public Model Ready Events; "Thinning", "Other Mechanical", "Clearcut", "Harvest", or "Reforestation"
- Include any of the following USGS GAP Land Cover classes; "Recently Logged Areas", "Harvested Forest - Grass/Forb Regeneration", "Harvested Forest-Shrub Regeneration", "Harvested Forest - Northwestern Conifer Regeneration", "Managed Tree Plantation", "Evergreen Plantation or Managed Pine", "Deciduous Plantations"
- Include either of the following USGS GAP Protected Areas Database classes where NLCD indicates "Forest" (41-43); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or Off Highway Vehicles (OHV) use" and "4 - no known mandate for protection"

Alaska

- Include either of the following USGS GAP Protected Areas Database classes where NLCD indicates "Forest" (41-43); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"

- Include any of the following USGS GAP Public Model Ready Events; "Thinning", "Other Mechanical", "Clearcut", "Harvest", or "Reforestation"
- AK LandFire EVT and GAP land cover do not have classes indicative of forest management

Hawaii

- Include the following LandFire EVT class; "Hawai'i Managed Tree Plantation"
- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- HI GAP land cover and USGS GAP Public Model Ready Events for HI do not have classes indicative of forest management

Puerto Rico

- Include the following GAP land cover classes; "Abandoned dry forest plantation", "Woody agriculture and plantations: Palm plantations"
- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- PR LandFire EVT is not available

Guam

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for Guam

Marianas

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Marianas

American Samoa

- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Marianas
- USGS GAP Protected Areas Database does not indicate areas indicative of forest management

Virgin Islands

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"

- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Virgin Islands

Managed Forests

Forested areas managed for timber extraction, forest tree plantations (excluding Christmas Trees). There is also a stand-alone Christmas Tree UDL, and a combine Forestry UDL which includes Managed Forest and Christmas Trees see above.

ConUS

- Include all the following LandFire Existing Vegetation Type (EVT) classes; "Recently Logged-Herb and Grass Cover", "Recently Logged-Shrub Cover", "Recently Logged-Tree Cover", "Managed Tree Plantation-Northern and Central Hardwood and Conifer Plantation Group", or "Managed Tree Plantation-Southeast Conifer and Hardwood Plantation Group"
- Include any of the following United States Geologic Survey (USGS) National Gap Analysis Program (GAP) Public Model Ready Events; "Thinning", "Other Mechanical", "Clearcut", "Harvest", or "Reforestation"
- Include any of the following USGS GAP Land Cover classes; "Recently Logged Areas", "Harvested Forest - Grass/Forb Regeneration", "Harvested Forest-Shrub Regeneration", "Harvested Forest - Northwestern Conifer Regeneration", "Managed Tree Plantation", "Evergreen Plantation or Managed Pine", "Deciduous Plantations"
- Include either of the following USGS GAP Protected Areas Database classes where NLCD indicates "Forest" (41-43); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or Off Highway Vehicles (OHV) use" and "4 - no known mandate for protection"

Alaska

- Include either of the following USGS GAP Protected Areas Database classes where NLCD indicates "Forest" (41-43); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- Include any of the following USGS GAP Public Model Ready Events; "Thinning", "Other Mechanical", "Clearcut", "Harvest", or "Reforestation"
- AK LandFire EVT and GAP land cover do not have classes indicative of forest management

Hawaii

- Include the following LandFire EVT class; "Hawai'i Managed Tree Plantation"
- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- HI GAP land cover and USGS GAP Public Model Ready Events for HI do not have classes indicative of forest management

Puerto Rico

- Include the following GAP land cover classes; "Abandoned dry forest plantation", "Woody agriculture and plantations: Palm plantations"

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- PR LandFire EVT is not available

Guam

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for Guam

Marianas

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Marianas

American Samoa

- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Marianas
- USGS GAP Protected Areas Database does not indicate areas indicative of forest management

Virgin Islands

- Include either of the following USGS GAP Protected Areas Database classes where CCAP indicates "Forest" (9-11); "3 - managed for multiple uses - subject to extractive (e.g., mining or logging) or OHV use" and "4 - no known mandate for protection"
- LandFire EVT, GAP land cover, and USGS GAP Public Model Ready Events are not available for the Marianas

Non-cultivated

Any land not captured by cultivated crops

ConUS

- Spatial represented as the inverse of all cultivated land as identified in USDA's Cropland Data Layer (2017).

Alaska (AK)

- Spatial represented as the inverse of the NLCD 2016 Cultivated Class (82)

Hawaii (HI)

- Spatial represented as the inverse of the CCAP 2011 Cultivated Class (6)

Puerto Rico (PR)

- Spatial represented as the inverse of the NLCD 2001 Cultivated Class (82)

Guam (GU)

- Spatial represented as the inverse of the CCAP 2011 Cultivated Class (6)

Marianas (CNMI)

- Spatial represented as the inverse of the CCAP 2004 Cultivated Class (6)

American Samoa (AS)

- Spatial represented as the inverse of the CCAP 2010 Cultivated Class (6)

Virgin Islands (VI)

- Spatial represented as the inverse of the CCAP 2012 Cultivated Class (6)

Nurseries

Non-agricultural Nurseries represent a land use that is not exclusive to any nationwide land cover class. Nurseries are mapped by using geocoded Dun and Bradstreet (D&B) business database addresses. Label uses that are covered by this UDL found on ornamentals, shrubs/vines, and non-food trees, grown in a non-agricultural setting (e.g., Retail Nurseries, Garden supply stores or retail horticultural location. This UDLs does not include label represented by agricultural nursery uses such as trees grown for food, tree plantations or transplanted trees, shrubs, and ornamentals. These agricultural nurseries are captured in the agricultural UDLs described above.

ConUS

- Using the Dun and Bradstreet business database, select all records with any SIC Codes starting with "018" (Horticultural Specialties) or "526" (Retail Nurseries, Lawn And Garden Supply Stores)
- Selected points are then buffered by their facility size attribute. Where facility size is absent, substitute the Census of Agriculture's average acreage by county, calculated using Nursery Totals. If a county's nursery acreages are undisclosed, then an average of all county averages is used. A circular buffer is applied, where radius is solved for using the areas previously described. In an effort to map production facilities only and not business offices, use the 'Location Type' attribute to categorize locations.

Alaska

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Hawaii

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Puerto Rico

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Guam

- No Dun and Bradstreet business data were available for Guam.

Marianas

- No Dun and Bradstreet business data were available for Marianas.

American Samoa

- No Dun and Bradstreet business data were available for American Samoa.

Virgin Islands

- Dun and Bradstreet business database was used in the same method as applied to ConUS.

Pasture/Rangeland (Grazing Areas)
The CDL and NLCD map a pasture class, but this is primarily grassland pastures. The grazing cattle land use is added to additional land cover types, such as forests, shrublands, wetlands, etc.

ConUS

- CDL (2013-2017) and NLCD 2016 pasture classes everywhere
- Excludes the cultivated agricultural grasses (captured in the alfalfa layer described above)
- Undeveloped NLCD classes within Bureau of Land Management (BLM) and United States Forest Service (USFS) grazing allotment boundaries
- Exclude NLCD developed, water, and cultivated

Alaska

- NLCD 2016 pasture class everywhere
- Undeveloped NLCD classes within BLM grazing allotment boundaries
- No USFS grazing allotment boundaries available for AK

Hawaii

- CCAP 2011 pasture class 7
- No BLM or USFS grazing allotment boundaries available for HI

Puerto Rico

- NLCD 2001 pasture class 81
- No BLM or USFS grazing allotment boundaries available for PR

Guam

- CCAP 2011 pasture class 7
- No BLM or USFS grazing allotment boundaries available for GU

Marianas

- CCAP 2004 pasture class 7

- No BLM or USFS grazing allotment boundaries available for CNMI

American Samoa

- CCAP 2010 pasture class 7
- No BLM or USFS grazing allotment boundaries available for AS

Virgin Islands

- CCAP 2012 pasture class 7
- No BLM or USFS grazing allotment boundaries available for VI

Right-of-Ways

Rights of way include areas that allow for passage on ground or property belong to another individual or government entity, these include areas such as roads, powerlines, railroad and pipelines. NLCD developed classes are sufficient for most scenarios. NLCD developed classes are insufficient in cases of rural minor roads, rural transmission lines, and rural pipelines.

ConUS

- All NLCD 2016 developed classes everywhere (21-24)
- For generating Euclidean distance for ConUS Right-of-Ways (ROW), NLCD Developed classes do not have Euclidean distance algorithms applied. NLCD Developed classes are included in the footprint as a zero value in the final Euclidean distance file. The other component ROW classes do have Euclidean distance algorithms applied.
- ESRI Railroads
- United States Census Bureau's Topologically Integrated Geographic Encoding and Referencing database (TIGER) transmission (MAF/TIGER Feature Class Code (MTFCC) code L4020) and pipeline (MTFCC code L4010) data
- Bonneville Power Administration's (BPA) Right-of-Way data
- NAVTEQ roads

Alaska

- All NLCD 2016 developed classes everywhere (21-24)
    - For generating Euclidean distance for Right-of-Ways (ROW), NLCD Developed classes do not have Euclidean distance algorithms applied. NLCD Developed classes are included in the footprint as a zero value in the final Euclidean distance file. The other component ROW classes do have Euclidean distance algorithms applied.
- ESRI Railroads
- United States Census Bureau's Topologically Integrated Geographic Encoding and Referencing database (TIGER) transmission (MAF/TIGER Feature Class Code (MTFCC) code L4020) and pipeline (MTFCC code L4010) data
- NAVTEQ roads

Hawaii

- CCAP 2011 developed classes everywhere (2-5)
- ESRI Railroads

- TIGER transmission (MTFCC code L4020) and pipeline (MTFCC code L4010) data
- NAVTEQ roads

Puerto Rico

- All NLCD 2001 developed classes everywhere (21-24)
- For generating Euclidean distance for Right-of-Ways (ROW), NLCD Developed classes do not have Euclidean distance algorithms applied. NLCD Developed classes are included in the footprint as a zero value in the final Euclidean distance file. The other component ROW classes do have Euclidean distance algorithms applied.
- ESRI Railroads
- United States Census Bureau's Topologically Integrated Geographic Encoding and Referencing database (TIGER) transmission (MAF/TIGER Feature Class Code (MTFCC) code L4020) and pipeline (MTFCC code L4010) data
- Navteq roads

Guam

- All CCAP 2011developed classes everywhere (2-5)
- No ESRI Railroads data available for Guam
- TIGER transmission (MTFCC code L4020) and pipeline (MTFCC code L4010) data
- No NAVTEQ roads data available for Guam

Marianas

- All CCAP 2004 developed classes everywhere (2-5)
- No ESRI Railroads data available for Marianas
- TIGER transmission (MTFCC code L4020) and pipeline (MTFCC code L4010) data
- No NAVTEQ roads data available for Marianas

American Samoa

- All CCAP developed classes everywhere (2-5)
- No ESRI Railroads data available for American Samoa
- No TIGER data available for American Samoa
- No NAVTEQ roads data available for American Samoa

Virgin Islands

- All CCAP developed classes everywhere (2-5)
- No ESRI Railroads data available for Virgin Islands
- No TIGER data available for Virgin Islands
- No NAVTEQ roads data available for Virgin Islands

APPENDIX C: Uncertainties and Conservative Assumptions Associated with the Overlap Analysis

The overlap analysis is based on the species locations provided by USFWS. Species range is defined as the geographical area where a species could be found in its lifetime. Produced and managed by the species experts in the agencies responsible for implementing the ESA, these data are the best available information for species range. EPA acknowledges that even though these are the best available data, there are several uncertainties. The range information is not sub-divided into additional qualifiers such as current/historical locations or temporal information to account for distribution variations relating to timing such as seasons. Without additional distribution information, EPA applies certain additional conservatisms: specifically, a uniform distribution within the range is assumed, meaning the species is assumed to be present in all sections of the range at all times of the year. This assumption is an additional conservatism because this distribution is unlikely to occur based a species life history.

Other commonly known and related sources of uncertainty for GIS data generally relate to accuracy and precision. Accuracy can be defined as how well information on a map matches the values in the real world. Precision relates to how well the description of the data used for mapping matches reality, based on closeness of repeated sets of measurements. The more precise the data, the more likely additional measurement or calculation will show the same result. Some sources of inaccuracy and imprecision in GIS data are obvious while others are difficult to identify. It is important to consider these sources of error as GIS software can make it appear that data are accurate and precise beyond the limits of the data. When conducting this spatial analysis to assess the relationship between the species range and agricultural location, EPA made conservative assumptions related to the accuracy and precision of the available data (e.g., using a 30 m resolution for the overlap process). These assumptions impact the uncertainty of the relationship, and generally overestimate the overlap between of species range and agricultural locations.

To address classification accuracy and positional accuracy of the agricultural GIS data used, EPA combines multiple years into a Use Data Layer (UDL) for each crop to represent anywhere the crop could be found. This is likely an overestimate of where a crop is found in any given year due to common agricultural practices such as rotation. Data resolution, or the smallest difference between features that could be recorded, is related to accuracy. The raster land cover data used to identify agricultural land, the Cropland data layer (CDL) produced by United States Department of Agriculture (USDA), has a resolution of 30 meters. A raster data set can be re-sampled into smaller increments, but this does not improve the resolution or accuracy of the dataset. For this reason, values cannot be established with a higher level of resolution than 30 meters, values that are not multiples of 30 cannot be determined (e.g., 30, 60, 90 are distance in the dataset; 50 is not).

Precision errors can be introduced when formatting data for processing. Formatting changes can include changes to scale, re-projections of data, and data format conversions (raster to vector or vice versa). Sources of errors that are not as obvious can include those originating from the initial measurements, digitizing of data and using different versions of a dataset. These types of precision error may introduce edge effect, or misaligned dataset when conducting the spatial analysis. Borders following the general

shape of the county boundaries but do not align exactly with range information used could be result of this type of precision error.

These uncertainties impact the relationship between the agricultural areas and species locations. EPA's spatial analysis makes conservative assumptions to err on the side of overestimating the potential for species exposure when assessing the relationship of the species range to agricultural land. EPA uses five years of crop information in constructing the UDLs representing the agricultural land, so that the UDLs include every location where the crop was grown during those five years. Due to normal agricultural practices (e.g., crop rotations), this is more land than expected in a given year. The relationship between the species and the agricultural land may be overestimated when the range is larger than the actual area occupied, and the additional area includes agricultural use or where edge effects were introduced. When considering the species location data, all areas may be occupied at the time the pesticide is used. County or state boundaries can be used as a conservative estimate for species range but species and natural habitats are not expected to follow man-made boundaries. When the species locations have not been refined beyond these man-made boundaries, underestimates of the relationship between species range and agricultural use can occur. While this underestimation is possible, EPA makes several conservative assumptions for agricultural land and species life history to account for this possibility. For agricultural land, use of the UDLs representing multiple years of agriculture, expands the agricultural footprint beyond what is expected in a given year. In addition to these assumptions, EPA uses the best available species location information from the species experts at USFWS and NMFS, minimizing this possibility.

References

Amos, J. J., C. M. Holmes, C. G. Hoogeweg, and S. A. Kay. 2010. Development of Datasets to Meet USEPA Threatened and Endangered Species Proximity to Potential Use Sites Data Requirements. Report Number: 437.01-Overview. Prepared by Waterborne Environmental, Inc. for the Generic Endangered Species Task Force.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witner. 1976. A land use and land cover classification system for use with remote sensor data. USGS Professional Paper. Vol. 964, 28 pp.

Boryan, C., Yang, Z., Mueller, R., & Craig, M. 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. Geocarto International, 26(5), 341-358. https://doi.org/10.1080/10106049.2011.562309

Baker, N. T., and Capel, P. D. 2011. Environmental factors that influence the location of crop agriculture in the conterminous United States: U.S. Geological Survey Scientific Investigations Report 2011–5108, 72 p.

Bolstad, P. 2005. GIS Fundamentals. 2nd edition. Eider Press, White Bear Lake, MN. 543 pp.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. Vol. 20, No. 1, pp. 37– 40.

Congalton, R. and K. Green. 2019. Assessing the Accuracy of Remotely Sensed Data Principles and Practices. Third Edition CRC Press, Boca Raton, FL 328pp.

Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C.

Johnson, D. M. 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. Agric. Syst., 114 (2013), pp. 95-105, 10.1016/j.agsy.2012.08.004

Johnston, C. A, 2013. Wetland losses due to row crop expansion in the dakota prairie pothole region Wetlands, 33, pp. 175-182, 10.1007/s13157-012-0365-x

Lark, T. J., R. M. Mueller, D. M. Johnson, and H. K. Gibbs. 2017. Measuring land-use andland-cover change using the U.S. department of agriculture's cropland data layer: Cautions and recommendations. International Journal of Applied Earth Observation and Geoinformation 62 (2017), pp 224-23. https://doi.org/10.1016/j.jag.2017.06.007

Shah, S. H., Y. Angel, R. Houborg, S. Ali, and M. F. McCabe. 2019. A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. Remote Sens. 11(8):920. https://doi.org/10.3390/rs11080920

Story, M. and R. Congalton. 1986. Accuracy assessment: A user' s perspective. Photogrammetric Engineering and Remote Sensing. Vol. 52, No. 3, pp. 397– 399.

Wright, C. K. and M. C. Wimberly, 2013.Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proc. Natl. Acad. Sci, 110 (10) pp. 4134-4139