**Metadata for Biocontrol Database**

Data discussed and interpreted in Karp *et al.* (2018) Crop pests and predators exhibit inconsistent responses to surrounding landscape composition. *PNAS.*

**Metadata file:**

The metadata file provides a description of each study in the database, as well relevant reference material. Columns include:

1. Study\_ID: The unique identifier for each included study.
2. Study\_Year: The years in which the study was conducted.
3. Country: The country in which the study was conducted.
4. Region: The region where the study was conducted. Authors were asked to describe a fine resolution descriptor of the region in which the study took place.
5. Land\_Use\_Change: Study authors reported whether or not their study area has undergone a significant amount of land-use change in recent years.
6. Covariates: Study authors reported whether there were any other factors that varied across their study sites that were important predictors of pest-control responses (*e.g.,* local diversity, soil moisture, organic status, *etc.*).
7. Sampling\_Design: Study authors were asked to describe how animal sampling was performed in their study.
8. Exclosure\_Treatments: Study authors were asked to state if their study involved exclosures, and, if so, to describe the treatments that were included and the number of replicates.
9. Function\_Methods: Study authors were asked to describe how enemy activity, pest activity, and crop yield data were collected.
10. Land.use.Source: The source of the land-use map that was used to calculate landscape composition around the study sites (see Karp *et al*. 2018 for more details).
11. Citations: References to any publications that have resulted from study as of May 2018.

**Sitedata file:**

The sitedata file provides a description of each study site and key covariates recorded at those sites.

1. Study\_ID: The unique identifier for each included study.
2. Site: A unique identifier for the study site (*i.e.,* one of the places were data was collected).
3. Farm: A unique identifier for the different farms where studies took places. Multiple sites could be located on the same farm.
4. Study\_Year: The year (or years) in which the site was sampled.
5. Crop\_Species: The crop species that were present at the study sites.
6. Annual\_Perennial: Whether the site hosted annual, perennial agriculture, or unknown.
7. Organic: Whether the site was certified organic (yes), conventional (no), or unknown.
8. Local\_Diversity: Whether the site was locally diverse (yes), not diverse (no), or unknown. Examples of local diversity include flowering strips, weedy margins, hedgerows, smaller field sizes, more diverse plantings, *etc*.
9. Insecticide\_Plot: This column indicates whether the study site was treated with insecticides (yes), not treated (no), or unknown. Insecticides could be applied through sprays, seed coating, planting transgenic crops, *etc.*
10. Insecticide\_Farm: This column indicates whether the farm was treated with insecticides (yes), not treated (no), or unknown. Insecticides could be applied through sprays, seed coating, planting transgenic crops, *etc.*
11. Confidence\_Insecticide\_Farm: This column details the percent confidence the authors had in their assessment that insecticides were or were not present on the farm, using a scale from 0% to 100%. Authors who wrote ‘unknown’ did not score their confidence.
12. Tilling: Whether the site was tilled (yes), not tilled (no), or unknown during the year of the study.
13. Site\_Covariate1: This space was provided if a different site-level covariate was found to be important in an author’s study system. Authors included information in this column at their discretion.
14. Site\_Covariate2: This space was provided if a second unique site-level covariate was found to be important in an author’s study system. Authors included information in this column at their discretion.

**Landusedata file:**

The landusedata file reports the landscape composition around each study site. Nine different land use classes (detailed below) were examined. From Karp *et al.* (2018):

“*We used a distance-weighting procedure to quantify landscape composition. First, we calculated the area (m2) of each land cover class in 20 concentric rings around each study site, with outer radii at 100 m intervals between 100 m and 2 km. Next, we used a Gaussian decay function to assign weights to each ring, where rings closer to the focal site were weighted more than those further away:*

*W= exp(-O2/(2\*d2) equation 1*

*where W is the weight, O is out the outer edge distance of the ring, and d is a decay rate that determines how rapidly weightings decline with distance. We used three decay rates (250, 1250, 1250) to later assess landscape composition at multiple spatial scales. We then used the weightings to calculate a weighted sum for the total area of each land cover class. Finally, we divided that number by the weighted sum of all land covers to ultimately obtain a proportional representation of each land-cover class around each study site.”*

Below, we report the Gaussian-decay-weighted estimates of landscape composition around each site.

1. Study\_ID: The unique identifier for each included study.
2. Site: A unique identifier for the study site (*i.e.,* one of the places were data was collected).
3. X: The x coordinate of the study site, in decimal degrees (WGS 1984, unprojected). Some coordinates were masked to protect grower anonymity.
4. Y: The y coordinate of the study site, in decimal degrees (WGS 1984, unprojected). Some coordinates were masked to protect grower anonymity.
5. LC1Gau250: The fraction of forest around the study site, using 250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
6. LC2Gau250: The fraction of grassland around the study site, using 250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
7. LC3Gau250: The fraction of shrubland around the study site, using 250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
8. LC4Gau250: The fraction of annual agriculture around the study site, using 250 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
9. LC5Gau250: The fraction of perennial agriculture around the study site, using 250 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
10. LC6Gau250: The fraction of other land uses around the study site, using 250 for the decay rate.
11. LC7Gau250: The fraction of urban areas around the study site, using 250 for the decay rate.
12. LC8Gau250: The fraction of ‘natural habitats’ around the study site, using 250 for the decay rate. This was used when the land use map did not sufficiently resolve forests grasslands, and/or shrublands. Otherwise, it is listed as 0.
13. LC9Gau250: The fraction of ‘agriculture’ around the study site, using 250 for the decay rate. This was used when the land use map did not sufficiently resolve annual and perennial agriculture. Otherwise, it is listed as 0.
14. LC1Gau750: The fraction of forest around the study site, using 750 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
15. LC2Gau750: The fraction of grassland around the study site, using 750 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
16. LC3Gau750: The fraction of shrubland around the study site, using 750 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
17. LC4Gau750: The fraction of annual agriculture around the study site, using 750 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
18. LC5Gau750: The fraction of perennial agriculture around the study site, using 750 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
19. LC6Gau750: The fraction of other land uses around the study site, using 750 for the decay rate.
20. LC7Gau750: The fraction of urban areas around the study site, using 750 for the decay rate.
21. LC8Gau750: The fraction of ‘natural habitats’ around the study site, using 750 for the decay rate. This was used when the land use map did not sufficiently resolve forests grasslands, and/or shrublands. Otherwise, it is listed as 0.
22. LC9Gau750: The fraction of ‘agriculture’ around the study site, using 750 for the decay rate. This was used when the land use map did not sufficiently resolve annual and perennial agriculture. Otherwise, it is listed as 0.
23. LC1Gau1250: The fraction of forest around the study site, using 1250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
24. LC2Gau1250: The fraction of grassland around the study site, using 1250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
25. LC3Gau1250: The fraction of shrubland around the study site, using 1250 for the decay rate. When the land-use map did not resolve differences between forests, grassland, and shrubland, ‘natural habitat’ was used as the category and all sites in the study received a 0 in this category.
26. LC4Gau1250: The fraction of annual agriculture around the study site, using 1250 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
27. LC5Gau1250: The fraction of perennial agriculture around the study site, using 1250 for the decay rate. When the land-use map did not resolve differences between annual and perennial agriculture, ‘agriculture’ was used as the category and all sites in the study received a 0 in this category.
28. LC6Gau1250: The fraction of other land uses around the study site, using 1250 for the decay rate.
29. LC7Gau1250: The fraction of urban areas around the study site, using 1250 for the decay rate.
30. LC8Gau1250: The fraction of ‘natural habitats’ around the study site, using 1250 for the decay rate. This was used when the land use map did not sufficiently resolve forests grasslands, and/or shrublands. Otherwise, it is listed as 0.
31. LC9Gau1250: The fraction of ‘agriculture’ around the study site, using 1250 for the decay rate. This was used when the land use map did not sufficiently resolve annual and perennial agriculture. Otherwise, it is listed as 0.

**Abundancedata file:**

The abundancedata details the abundances of all pests and/or natural enemies recorded at the study sites.

1. Study\_ID: The unique identifier for each included study.
2. Site: A unique identifier for the study site (*i.e.,* one of the places were data was collected).
3. Study\_Year: The year (or years) in which the site was sampled.
4. Sampling\_Date: A unique identifier that differs for each site when it was visited on a different day within the same year.
5. Species: The scientific name of the species that was surveyed.
6. Family: The family of the species that was surveyed.
7. Order: The order of the species that was surveyed.
8. Functional\_Group: The functional group of the species that was surveyed. Possibilities include: dominant pest (economically damaging), secondary pest (not strongly damaging but still herbivorous), dominant enemy (a predator or parasitoid that is known to control a dominant pest), parasitoid, or predator.
9. Abundance: The recorded abundance of the species. This value differs across studies, with some reporting the number of individual recorded and others reporting average abundances/counts across multiple surveys.
10. Number\_Censuses: The number of sub-samples that were summarized in the abundance column (*e.g.,* three pan traps averaged per plot would be recorded as 3).
11. Sampling\_Method: The method used to sample organisms (*e.g.,* pan traps, pitfall traps, mist nets, sticky cards, *etc.*)
12. Abundance\_Duration: The number of days that the sampling lasted (*e.g.,* how many days the sticky trap was left in the field).
13. Notes: Any notes that the study’s author felt important to include.

**Activitydata file:**

The activitydata file details the activity of pests (*e.g.,* crop damage) and enemies (*e.g.,* percent pests parasitized) that was recorded at the study sites.

1. Study\_ID: The unique identifier for each included study.
2. Site: A unique identifier for the study site (*i.e.,* one of the places were data was collected).
3. Study\_Year: The year (or years) in which the site was sampled.
4. Sampling\_Date: A unique identifier that differs for each site when it was visited on a different day within the same year.
5. Function\_Type: The broad category of data being collected. Can be one of the categories:
   1. Cage experiment (change in infestation): pest abundance recorded within an exclosure experiment
   2. Cage experiment (change in crop damage): crop damage recorded within an exclosure experiment
   3. Cage experiment (change in crop yield): crop yield recorded within an exclosure experiment
   4. Sentinel experiments (enemy activity): the percent predation or parasitism observed after placing sentinel pests in the field or, for parasitism, the fraction of parasitized pests observed naturally in the field.
   5. Pest damage: crop damage incurred by pest feeding.
6. Pest: The name of the pest species which is being studied (*i.e.,* the pest that is being consumed or parasitized or the pest that is exacting damage on the crop). If the pest is not differentiated, then ‘all’ is written.
7. Exclosure\_Treatment: If an exclosure is present, this column indicates whether the data reported are from the ‘open’ treatment (allowing predators to be present) or the ‘closed’ treatment (excluding predators).
8. Initial\_Density: If the study involved exclosures, this column indicates the level of function data before the exclosures were erected(*e.g.,* the starting abundances of the pests).
9. Function\_Data: The value of the measured function in the exclosure treatment and/or at the site.
10. Unit: The unit of measurement for the function data (*e.g.,* percentage of pests consumed).
11. Number\_Censuses: The number of sub-samples that were summarized in activity column (*e.g.,* three cage traps averaged per site would be recorded as 3).
12. Activity\_Duration: The number of days that the sampling lasted (*e.g.,* how many days the sentinel pests was left in the field).
13. Notes: Any notes that the study’s author felt important to include.

**Yielddata file:**

The yielddata file details the crop yields that were recorded at the study sites.

1. Study\_ID: The unique identifier for each included study.
2. Site: A unique identifier for the study site (*i.e.,* one of the places were data was collected).
3. Study\_Year: The year (or years) in which the site was sampled.
4. Sampling\_Date: A unique identifier that differs for each site when it was visited on a different day within the same year.
5. Function\_Data: The reported value of crop yield at a given site.
6. Unit: The unit of measurement for the yield data (*e.g.,* kg/ha).
7. Measure\_Type: A variable indicating whether the yield was recorded as the total yield of the plant (*e.g.,* total biomass)or within-plant-yield.
8. Sample\_Size: The number of plants or sampled area that went in to calculating the yield estimate.
9. Marketable\_or\_Total: A variable indicated whether the estimate of yield reported corresponded to the total yield of the crop or just the fraction that was marketable (*e.g.,* due to quality concerns).
10. Notes: Any notes that the study’s author felt important to include.
11. CropType: The crop species for which the yield estimate was calculated.
12. WeeksSincePlanting: The number of weeks since planting when the yield estimate obtained in the field.