

Pesticides and Bee Biodiversity at the Golden Coast

Joshua Paul Cohen

2024-12-19

Table of contents

Introduction	2
Key Highlights	4
Repository Link	4
Data	4
Big Bee: Extending Anthophila Research Through Image and Trait Digitization	4
California DPR Annual Pesticide Summaries	5
California Public Land Survey System	5
USGS 2023 Land Cover (CONUS)	5
Analysis Plan	6
Analysis	7
Final Figure	16
Discussion	19
Reflection	19
Possible Next Steps	20
Citations	21
In Text	21
Data	22

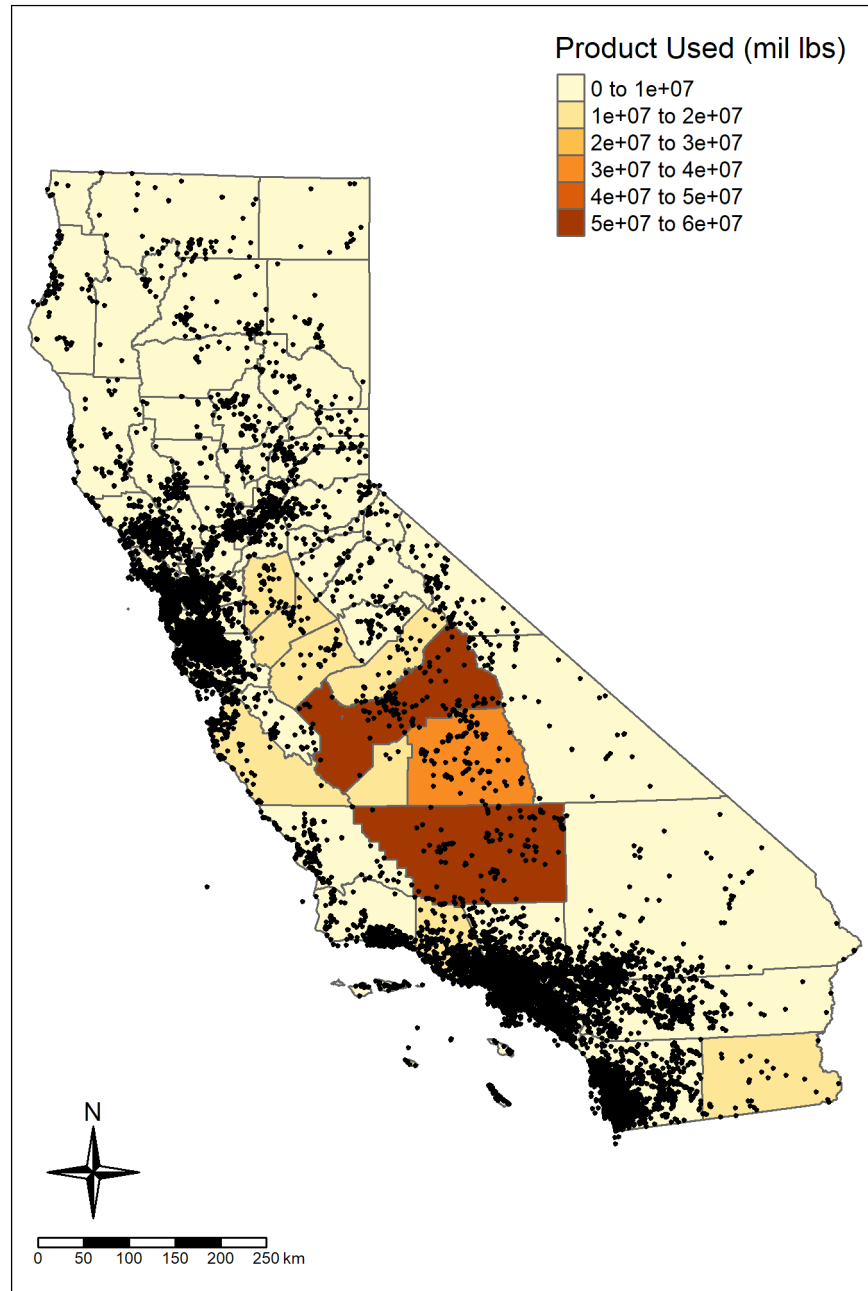
Introduction

Bees and other pollinators are extremely valuable for the ecosystem services they provide. Yet, throughout the past few decades, they have declined in both population and biodiversity (Goulson et al, 2015). Attributed factors are invasive parasites (Meeus et al, 2011), climate change (Soroye, Newbold, & Kerr, 2020), habitat destruction (and fragmentation (Aizen & Feinsinger, 1994), and pesticide use (Arena & Sgolastra, 2014).

Several factor's make California an ideal microcosm to analyze the effect of pesticides on bee biodiversity.

- The state's 1989 Food Safety Act requires uniquely high standards for reporting (California DPR, 2023). Pesticide use datasets out of CA are therefore large and highly detailed.
- CA is size means it has many different ecosystems contained within the same political boundaries.
- CA is a agricultural hotspot in the country, meaning there are areas of intensive pesticide use.

Therefore in this analysis, we will be analyzing pesticide use. Later, we introduce another variable, land cover, to see if there is a significant effect.



Pesticide use map by county, with bee species points overlaid. Most points are concentrated within Bay Area, Los Angeles, and San Diego.

Key Highlights

- Creation of two error models
 - A two variable model between biodiversity and pesticide use.
 - An extrapolation of the previous model with an added predictor of land cover.
- Diagnosing spatial autocorrelation in our models by running a lag model on the residuals and finding Moran's I.
- A hypothesis test for the affect of land cover on the model where we rejected the null.

Repository Link

All code for this project [can be accessed here](#).

Data

Big Bee: Extending Anthophila Research Through Image and Trait Digitization

- Description
 - A large data base containing 4 million+ records of bee occurrence data for 5000+ unique species, along with attribute data.
 - Compilation of multiple databases.
 - Format: csv with latitude/longitude columns.
 - Contains observations from around the world, and from as early as the 19th century.
 - Led by Katja Seltmann, a researcher associated with the Cheadle Center for Biodiversity and Restoration at UCSB.
 - Overarching goal is to compile bee attribute data (body length, flight range, etc) for as many species as possible into one place.
- Limitations
 - Spatially biased: most observations are located near areas of high population
 - Dirty data: A compilation approach to sampling leads to inconsistent data structure with many NAs.
 - No absence data.
 - Includes citizen science data, increasing chance of possible misidentification.
- Subset in this analysis
 - Data from the past 5 years in all of California.
- Access

- Data is downloadable from [the map section of the Big Bee Library](#).
- Must choose a subset to download, can subset by location, taxa, time, observer/collector, and more.

California DPR Annual Pesticide Summaries

- Description
 - In accordance with the Food Safety Act of 1989.
 - Contains data for products, active ingredients, amounts, locations, etc.
 - Format: comma delineated .txt files.
- Limitations
 - Does not contain spatial information, only analyzable spatially through joining.
- Subset in this analysis
 - All available data from 2017.
- Access Data is downloadable from [California DPR's Pesticide Use Reporting page](#).

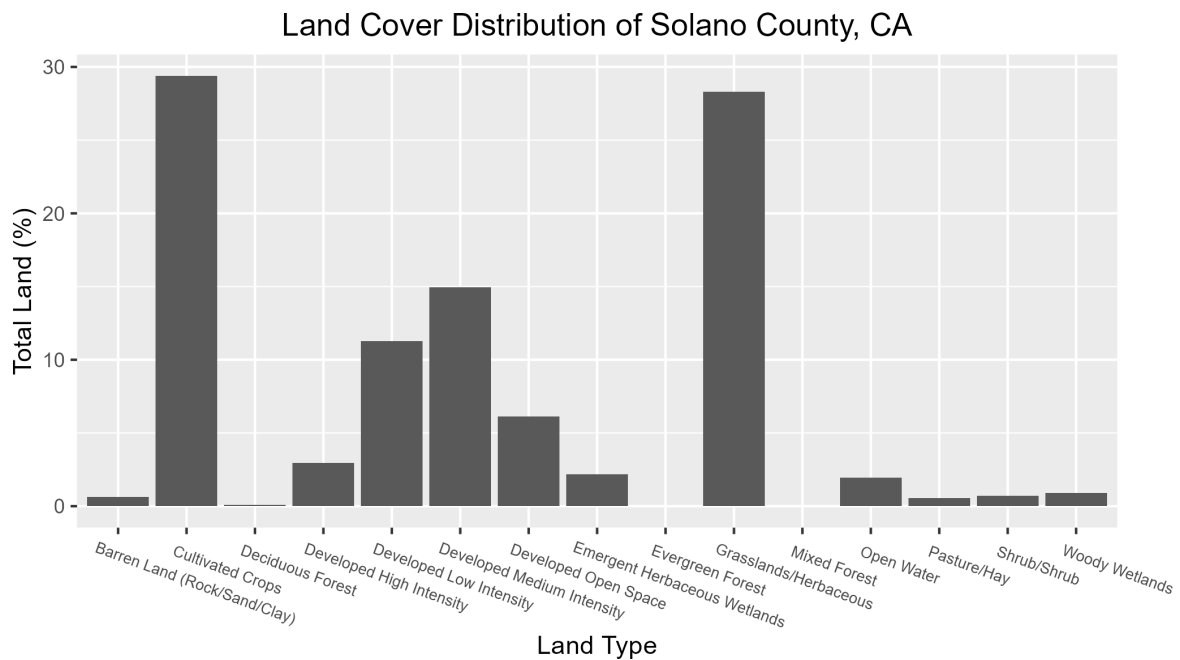
California Public Land Survey System

- Description
 - Shapefile associated with the pesticide use data
 - Contains polygons for all CA townships.
- Limitations
 - Pesticide data does not contains a joinable column for township, only joinable by county.
- Access
 - See pesticide use data.

USGS 2023 Land Cover (CONUS)

- Description
 - Land cover classification types.
 - 16 different classification types for water, development, forest/grassland, marsh, barren land, and agriculture.
 - CONUS stands for Conterminous United States.

- Limitations
 - Extremely large, and designed for a county wide scale. Needed to be aggregated to be worked with on a state wide scale
 - Aggregation caused distortion in classification values.
- Access
 - Data is downloadable from [it's page on the official USGS website](#).
 - It is possible to subset the data by uploading a county shapefile [in the MRLC Viewer](#), then download each county individually, but this was not done due to time restrictions.



A demonstration of a small subset of the data within the USGS Land Cover Dataset. Crop-land and grassland are common land types in Solano County, with development also being notable.

Analysis Plan

1. Initial data processing.
2. Define and create linear model.

3. Diagnose spatial autocorrelation, and create error model if evidence for autocorrelation is found.
4. Test for other correlating variables if initial model is not significant.
5. Create a new parallel slopes model with new variable.
6. Repeat steps 2 and 3 with new model.
7. Fitting a loess model.
8. Running a hypothesis test.

Analysis

1. Join data

- The pesticide data was summarized by county.
- Bee occurrence data was subsetting to just records ID'd down to species.
- Pesticide data and CA County shapefile is joinable by `county_cd` (county code).
 - The shapefile contains polygons for every township district in the state, but there is no column in the pesticide data to join on that scale, so we must aggregate up to county.
- Shapefile and bee occurrence data are joinable by spatial relationship.

```
# group by county and join population
counties <- townships %>%
  clean_names() %>%
  group_by(county_cd) %>%
  summarize(geometry = st_union(geometry)) %>%
  inner_join(pop_mean, by = join_by('county_cd'))

# get pesticide sum by county
pesticides_sum <- pesticides %>%
  select(lbs_prd_used, county_cd) %>%
  group_by(county_cd) %>%
  summarize(lbs_prd_used = sum(lbs_prd_used, na.rm = TRUE))

# join by county
pest_by_county <- left_join(counties, pesticides_sum, by = join_by(county_cd))

# select for rows with ids down to species
bee_species <- bees %>%
  filter(taxonRank == 'Species') %>%
  select('id', 'scientificName')
```

```
# get species richness, by county
biodiversity <- st_join(pest_by_county, bee_species) %>%
  group_by(county_name) %>%
  summarize(lbs_prd_used = first(lbs_prd_used),
            sp_richness = n_distinct(scientificName)/n())
```

2. Define linear model

Initially, I tested if a linear model was appropriate. It is written as:

$$\text{Species Richness} = \beta_0 + \beta_1 \text{Pesticide Amount}$$

I then created it in R:

```
biodiv_lm <- lm(sp_richness ~ kg_prd_used, biodiversity)
```

```
Call:
lm(formula = sp_richness ~ kg_prd_used, data = biodiversity)

Residuals:
    Min       1Q   Median       3Q      Max
-0.18217 -0.14059 -0.04673  0.07545  0.70276

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.979e-01  2.603e-02   7.602 3.51e-10 ***
kg_prd_used -7.221e-09  4.744e-09  -1.522   0.134
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1697 on 56 degrees of freedom
Multiple R-squared:  0.03972,    Adjusted R-squared:  0.02257
F-statistic: 2.316 on 1 and 56 DF,  p-value: 0.1336
```

Species richness and pesticide use are not coorelated upon initial inspection, as the R2 is very low and p-value is above 0.10.

3. Diagnose spatial autocorrelation.

Spatial data often violates the homoscedasticity assumption of Only Least Squares (OLS), skewing the results when creating a basic linear model, due to Tobler's first law of geography.

```
# get linear model
biodiv_lm <- lm(sp_richness ~ kg_prd_used, biodiversity)

# attach resid to dataset
biodiversity <- biodiversity %>% mutate(sp_pest_resid = resid(biodiv_lm))

# get centroids of each county polygon
bio_center <- st_centroid(biodiversity)

# try and calculate spatial autocorrelation at 50 km
bee_nb <- dnearneigh(bio_center, d1 = 0, d2 = 100000)

# assign weights to each neighboring object
bee_lw <- nb2listw(bee_nb, style = 'W', zero.policy = TRUE)

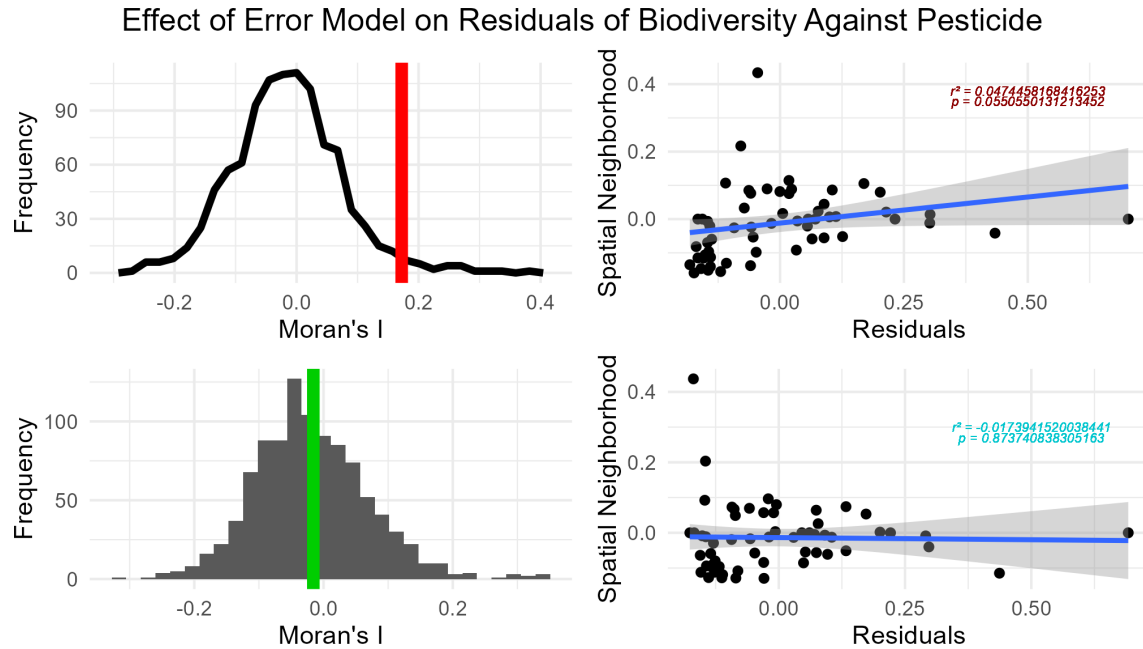
# get moran I
bee_resid_moran <- moran(biodiversity$sp_pest_resid,
                        listw = bee_lw,
                        n = length(bee_nb),
                        S0 = Szero(bee_lw))

# run 999 simulations
bee_mc <- moran.mc(biodiversity$sp_pest_resid, bee_lw, nsim = 999)

# running a lag model
lag_mod <- lag.listw(bee_lw, biodiversity$sp_pest_resid)

# making lag model table
lag_mod_tib <- tibble(resids = biodiversity$sp_pest_resid, neighborhood = lag_mod)

# lag mod lm
lag_lm <- lm(neighborhood ~ resids, data = lag_mod_tib)
```



Demonstration of spatial autocorrelation. On top, Moran's I falls outside the range of the null distribution, while the residuals have a slight positive correlation with the spatial neighborhood. The bottom figures are plotted with a error model, remediating the autocorrelation.

Because spatial autocorrelation is present, we create an error model.

```
bio_err <- errorsarlm(sp_richness ~ kg_prd_used,
  data = biodiversity,
  listw = bee_lw,
  na.action = na.exclude,
  zero.policy = TRUE)

summary(bio_err)
```

We had run a lag model to correct for the autocorrelation. A lag model is a model in which the variables are offset step by step from one another, and that offset can be plotted on a figure with the lag on the x-axis, and the neighborhood on the y-axis. Variables that have autocorrelation between variables will hang for a while before hovering around the mean, whereas variables with no spatial autocorrelation nearly immediately begin hovering around the mean.

```
# add model and model residuals column to lag tibble
if(exists('error_mod') == FALSE){
lag_mod_tib <- lag_mod_tib %>% add_column(error_mod = fitted(bio_err), error_resid = resid(b
```

```

}

# running a lag model with the error model
lag_mod_err <- lag.listw(bee_lw, lag_mod_tib$error_resid)

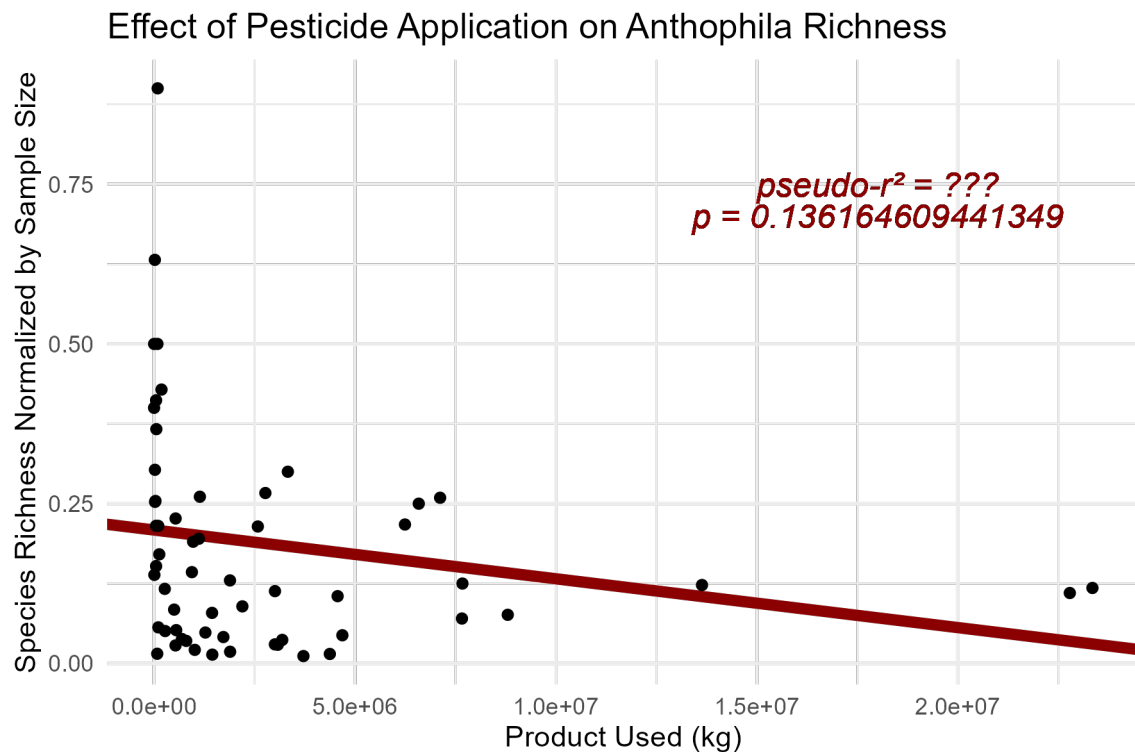
# get moran I for error model
err_moran <- moran(lag_mod_tib$error_resid, listw = bee_lw, n = length(bee_nb), S0 = Szero(b

# run simulation with error model
resid_mc <- moran.mc(lag_mod_tib$error_resid, bee_lw, nsim = 999)

# add lag for error to tibble
if ('neighborhood_err' %in% names(lag_mod_tib) == FALSE){
lag_mod_tib <- lag_mod_tib %>% add_column(neighborhood_err = lag_mod_err)
}

# run an lm to get p for errors neighborhood ~ resid
err_res_lm <- lm(lag_mod_tib$neighborhood_err ~ lag_mod_tib$error_resid)

```



Pesticide used has been plotted against Biodiversity in a simple two variable linear model. The

variables are not significant, but there is a slight negative slope.

4. Test for other correlating variables.

Because the relationship between pesticide use and biodiversity was not found to be significant, I tested other variables. These variables are:

Habitat Fragmentation

```
lm(formula = sp_richness ~ habitat_frag, data = biodiversity)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.12473    0.05341   2.335  0.0231 *
habitat_frag  0.01216    0.01118   1.087  0.2815
Residual standard error: 0.1714 on 56 degrees of freedom
Multiple R-squared:  0.02068,    Adjusted R-squared:  0.003192
F-statistic: 1.183 on 1 and 56 DF,  p-value: 0.2815
```

Population Density

```
lm(formula = sp_richness ~ pop_density, data = biodiversity)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.024e-01  2.322e-02   8.715 5.22e-12 ***
pop_density -1.285e-04  4.692e-05  -2.739  0.00826 **
Residual standard error: 0.1626 on 56 degrees of freedom
Multiple R-squared:  0.1181,    Adjusted R-squared:  0.1024
F-statistic:  7.5 on 1 and 56 DF,  p-value: 0.008257
```

Land Use

```
lm(formula = sp_richness ~ land_use, data = biodiversity)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10572    0.02570   4.114 0.000129 ***
land_use      0.17324    0.03995   4.336  6.1e-05 ***
Residual standard error: 0.1499 on 56 degrees of freedom
Multiple R-squared:  0.2514,    Adjusted R-squared:  0.238
F-statistic: 18.8 on 1 and 56 DF,  p-value: 6.101e-05
```

5. Add new variable to model.

Because land use was the most significantly correlated variable. I added it to the model:

```
lm(formula = sp_richness ~ kg_prd_used + land_use, data = biodiversity)
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.246e-01  2.845e-02   4.380 5.38e-05 ***
kg_prd_used  -6.149e-09  4.153e-09  -1.481  0.144
land_use      1.697e-01  3.961e-02   4.285 7.41e-05 ***
Residual standard error: 0.1483 on 55 degrees of freedom
Multiple R-squared:  0.2801,    Adjusted R-squared:  0.2539
F-statistic: 10.7 on 2 and 55 DF,  p-value: 0.000119
```

Land use as a variable is related to habitat fragmentation. In this analysis, we added it as a binary variable, separating the land types by habitat and non-habitat.

Our new model is:

$$\text{Species Richness} = \beta_0 + \beta_1 \text{Pesticide Amount} + \beta_2 \text{Land Use}$$

```
Call:
lm(formula = sp_richness ~ kg_prd_used + land_use, data = biodiversity)

Residuals:
    Min       1Q   Median       3Q      Max
-0.25262 -0.08129 -0.03860  0.06577  0.60623

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.246e-01  2.845e-02   4.380 5.38e-05 ***
kg_prd_used  -6.149e-09  4.153e-09  -1.481  0.144
land_use      1.697e-01  3.961e-02   4.285 7.41e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1483 on 55 degrees of freedom
Multiple R-squared:  0.2801,    Adjusted R-squared:  0.2539
F-statistic: 10.7 on 2 and 55 DF,  p-value: 0.000119
```

Adding land use to the model, substantially increased the r^2 and made the p-value significant at $p < 0.01$.

6. Spatial autocorrelation diagnosis/error model with added variable.

The code and figures are very similar from before.

```
# new lm w land cover
land_use_lm <- summary(lm(sp_richness ~ kg_prd_used + land_use, biodiversity))

# attaching resid of new lm to df
biodiversity <- biodiversity %>% mutate(sp_pest_land_resid = resid(land_use_lm))

# get moran I again for the model with land cover
bee_resid_moran <- moran(biodiversity$sp_pest_land_resid,
                        listw = bee_lw,
                        n = length(bee_nb),
                        S0 = Szero(bee_lw))

# run 999 simulations
bee_mc <- moran.mc(biodiversity$sp_pest_land_resid, bee_lw, nsim = 999)

# running a lag model
lag_mod <- lag.listw(bee_lw, biodiversity$sp_pest_land_resid)

# making lag model table
lag_mod_tib <- tibble(resids = biodiversity$sp_pest_land_resid, neighborhood = lag_mod)

# lag mod lm
lag_lm <- lm(neighborhood ~ resids, data = lag_mod_tib)
```

The neighborhood weighting from the previous error model is already suitable for this one so there's no need to do it again.

There was possibly some autocorrelation. While the r^2 of the residuals is close to 0, the Moran's I was still somewhat off from the null dist.

Making the error model and running the lag model:

```
bio_land_err <- errorsarlm(sp_richness ~ kg_prd_used + land_use,
                          data = biodiversity,
                          listw = bee_lw,
                          na.action = na.exclude,
                          zero.policy = TRUE)
```

```

summary(bio_land_err)

# add model and model residuals column to lag tibble
if(exists('error_mod_land') == FALSE){
lag_mod_tib <- lag_mod_tib %>% add_column(error_mod_land = fitted(bio_land_err), error_resid_
}

# running a lag model with the error model
lag_mod_err <- lag.listw(bee_lw, lag_mod_tib$error_resid_land)

# get moran I for error model
err_moran <- moran(lag_mod_tib$error_resid_land, listw = bee_lw, n = length(bee_nb), S0 = Sz

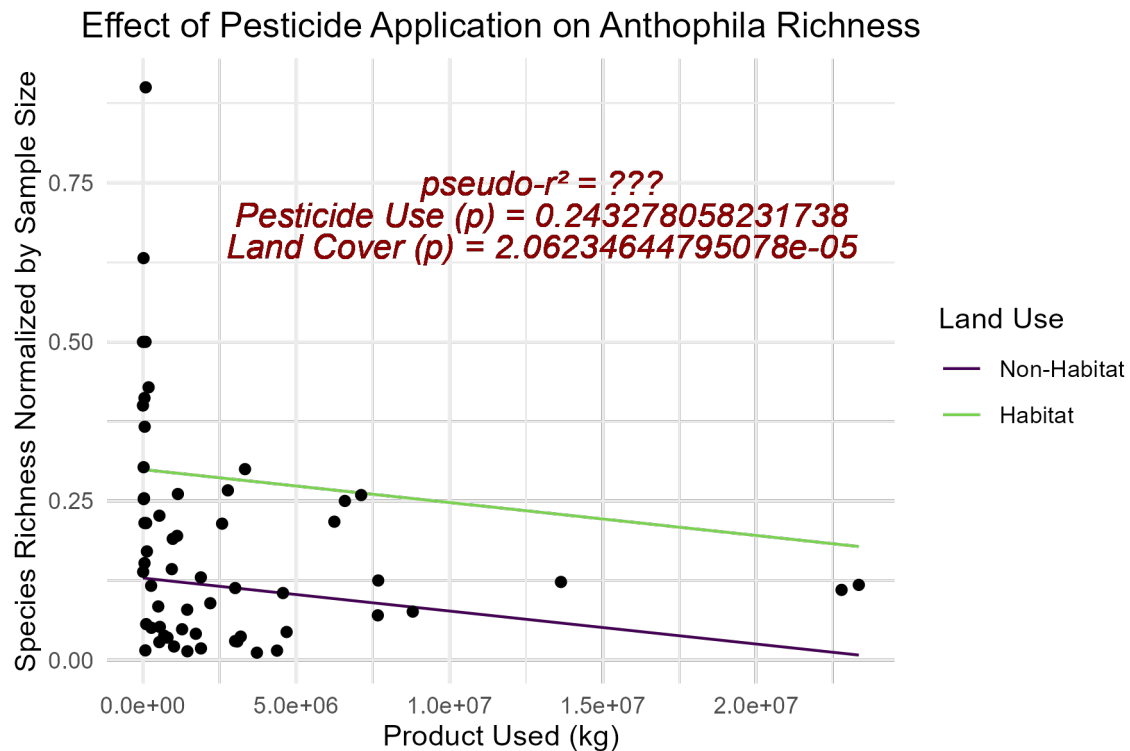
# run simulation with error model
resid_mc <- moran.mc(lag_mod_tib$error_resid_land, bee_lw, nsim = 999)

# add lag for error to tibble
if ('neighborhood_err_land' %in% names(lag_mod_tib) == FALSE){
lag_mod_tib <- lag_mod_tib %>% add_column(neighborhood_err_land = lag_mod_err)
}

# run an lm to get p for errors neighborhood ~ resid
err_res_lm <- lm(lag_mod_tib$neighborhood_err_land ~ lag_mod_tib$error_resid_land)

```

Final Figure

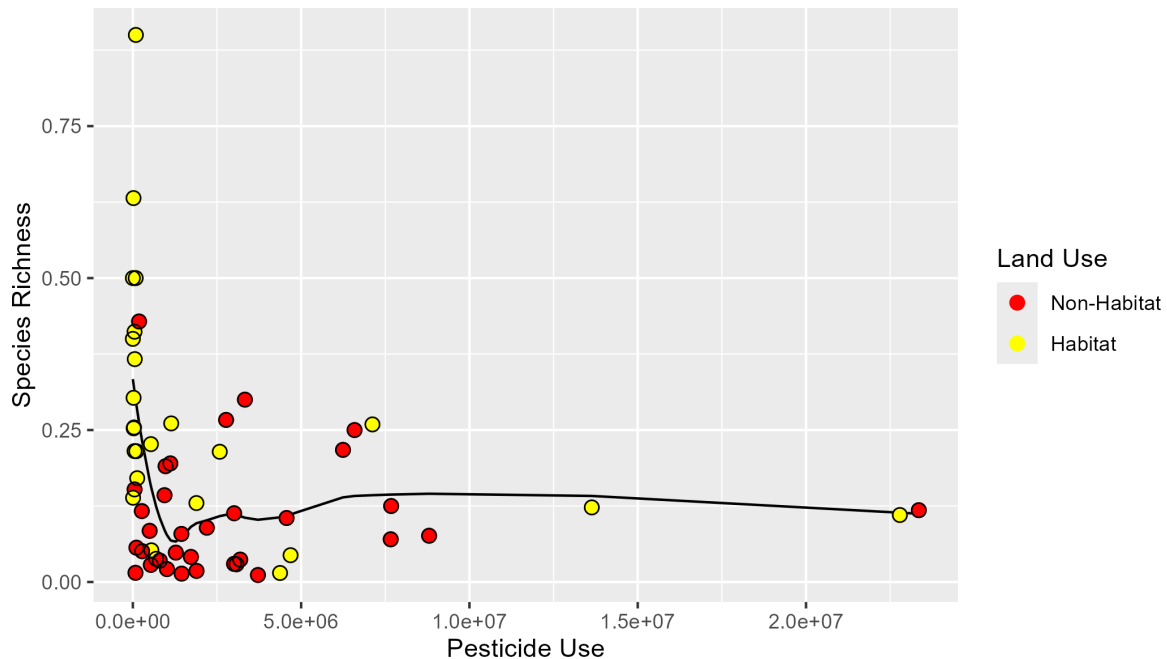


A parallel slopes model for the effect of pesticide use and land cover on species richness. There is more biodiversity in habitat land types such as grasslands and forests than there are in non-habitat land types such as development. While the pesticide use variable was not significant, the land cover variable was. Species richness has been normalized by sample size.

While our relationship between pesticide use and biodiversity still was not significant, the relationship between biodiversity and land cover was. In this parallel slopes model, the difference between bee habitat land and non bee habitat land is distinct.

7. Fitting a loess model.

Because the points for the model appear to follow an exponential decay curve, I decided a loess model potentially could be appropriate.



Loess model for our parallel slopes model between biodiversity, pesticide use, and land use.

8. Running a hypothesis test.

Despite the result, I will still see if it was a product of chance or not.

1. Declare the hypothesis

- H_0 = There is no significant effect of habitat on the relationship between biodiversity and pesticide use.
- H_A = There is a significant effect of habitat on the relationship between biodiversity and pesticide use.

2. Obtain t-statistic

- Our t-statistic is...
 - The difference in y-intercept for the model between habitat and non-habitat land.

```
# separate land use values
biodiversity_hab <- biodiversity %>%
  filter(land_use == 1)
biodiversity_non_hab <- biodiversity %>%
  filter(land_use == 0)
```

```

# make linear models to get y ints from
hab_error <- lm(sp_richness ~ kg_prd_used, biodiversity_hab)
non_hab_error <- lm(sp_richness ~ kg_prd_used, biodiversity_non_hab)

# get y intercepts
t_st_hab <- coef(summary(hab_error))[1,1]
t_st_non_hab <- coef(summary(non_hab_error))[1,1]

# get difference
diff <- t_st_hab - t_st_non_hab

```

3. Run a randomization test

```

biodiversity_null <- replicate(1000, {

# shuffle habitat in biodiversity
biodiversity_shuffle <- biodiversity %>%
  mutate(land_use = rbinom(nrow(biodiversity), 1, 0.5))

# separate land use values
biodiversity_hab <- biodiversity_shuffle %>%
  filter(land_use == 1)
biodiversity_non_hab <- biodiversity_shuffle %>%
  filter(land_use == 0)

# make error models to get y ints from
hab_error <- lm(sp_richness ~ kg_prd_used, biodiversity_hab)

non_hab_error <- lm(sp_richness ~ kg_prd_used, biodiversity_non_hab)

# get y intercepts
t_st_hab <- coef(summary(hab_error))[1,1]
t_st_non_hab <- coef(summary(non_hab_error))[1,1]

# get difference
diff <- t_st_hab - t_st_non_hab

})

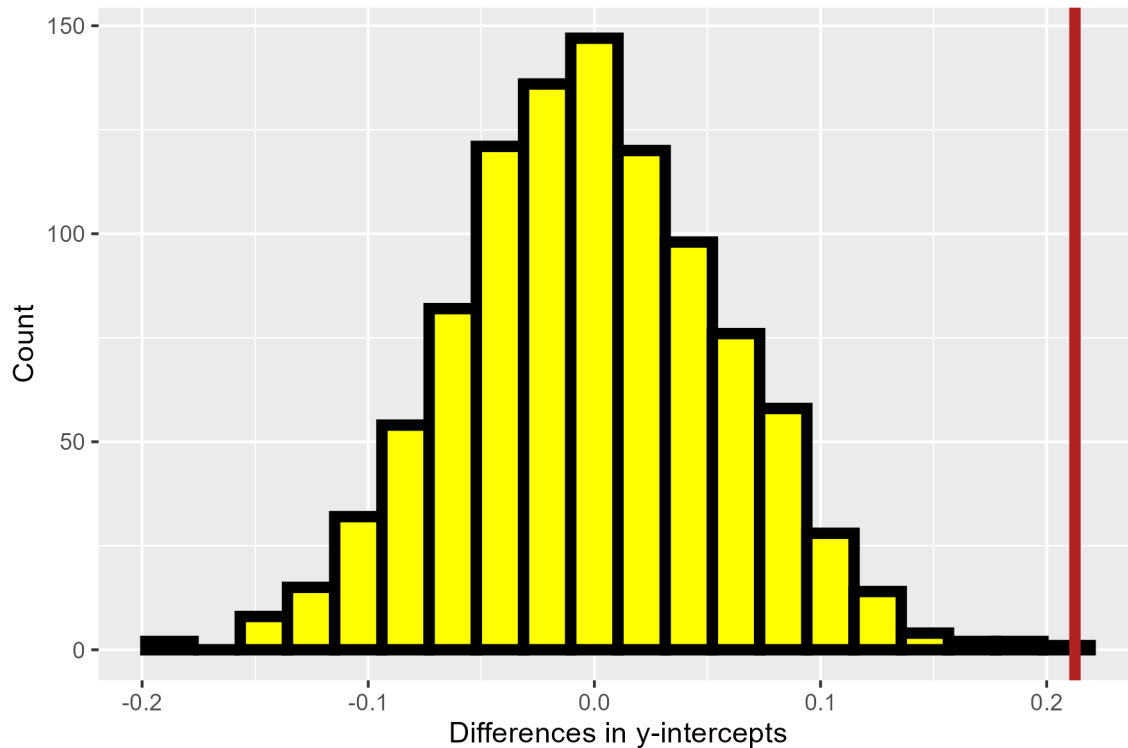
```

4. Obtain p-value

```

p_value <- sum(biodiversity_null > diff) /length(biodiversity_null)

```



A histogram of a normal distribution for the randomization test. The difference in y intercept was obtained 1000 times when randomly shuffling the land cover variable.

5. Accept or reject the null hypothesis

Because the p-value is significantly below $p = 0.05$, we can reject the null.

This is possibly an erroneous result, as it's extremely unlikely for our t-statistic to be so far outside the null distribution.

Discussion

Reflection

While I hypothesized that we would see a significant effect between biodiversity and pesticide use, we didn't. There are several possible reasons for this.

1. Complexity of ecological forces

- While it remains true that bee populations are in decline, it's also a somewhat complicated topic. For one, not all bee species are equally vulnerable. Bumblebee's longer flight ranges and bigger bodies allow them to tough out the dredge of a long flight from one isolate habitat patch to another and the toxicity of pesticides better than honeybees do (Greenleaf et al, 2007). Furthermore, bees have shown to have a significant amount of phenotypic plasticity, in which they are able to modify their behavior according to their environment (Corona, Libbrecht, & Wheeler, 2016). Furthermore, a significant amount of invasive pressure comes from other bees - the European Honeybee.

2. Analysis biases

- Poor data precision
 - Disappointingly, the pesticide data, in spite of all it's detail, did not have a column on it in order to join with the shapefile that came with it by township. This possibly affected the results of this analysis, as all the data, even if it was originally in great detail, had to be aggregated up to the county level.
 - The hypothesis test was notable as it was likely flawed. For one, I was not able to use an error model due to a strange error in the hypothesis test. Secondly, my p-statistic was suspiciously low.

3. Data biases

- Uneven distribution of data points
 - Big Bee, a database that contains data from a multitude of sources is unlikely to have it's data be distributed evenly in space for if no other reason that data is more likely to be collected where people already are. It has many different partner and sources but it also collects research grade iNaturalist observations meaning it's possible there will be less observations out in the middle of the desert and more in a popular park and picnic spot.

Possible Next Steps

Bee ecology and decline is a complex topic and there is still much left to learn. Ways I could continue this analysis include...

1. Refine this analysis and correct possible mistakes

- There are a multitude of ways this analysis could be improved.
 - How do we visualize the pesticide data with more precision? It's possible there is something I missed, and the CA DPR may be able to help. Being about to accomplish this could lead to more interesting results than the ones obtained here.
 - Another possibility is to apply a logistic regression model to the land_cover variable. Because it's the variable that appears to be most significantly correlated, it could be investigated further. Because it is such a high resolution dataset with many different land types, it may warrant it's own analysis.

2. Investigate other variables

- Variables such as the influence of parasites and climate change are things that weren't at all included in this analysis. Invasive parasites have been cited as being a notable contributor to colony collapse disorder (Meeus et al, 2011), and climate change continues to have it's own negative effects (Soroye, Newbold, & Kerr, 2020).

3. Possible interactions with non-bee factors

- The environment is confusing and wild, and nothing will ever make sense to us but that makes it all the more beautiful. Of course, this won't stop us from trying. Is sea level rise affecting a particular taxa's favorite flower? Is there some stochastic connection between bee activity and monarch migration? These are things we can only know if they're investigated.

Citations

In Text

Goulson, D., Nicholls, E., Botías, C., & Rotheray, E. L. (2015). Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science*, 347(6229), 1255957.

California Department of Pesticide Regulation. (2023, January). PESTICIDE USE ANNUAL REPORT. https://www.cdpr.ca.gov/docs/pur/pur_history_background.pdf

Meeus, I., Brown, M. J. F., de Graaf, D. C., & Smagghe, G. (2011). Effects of invasive parasites on bumble bee declines. *Conservation Biology*, 25(4), 662-671. <https://doi.org/10.1111/j.1523-1739.2011.01707.x>

Soroye, P., Newbold, T., & Kerr, J. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science*, 367(6478), 685-688.

- Aizen, M. A., & Feinsinger, P. (1994). Habitat fragmentation, native insect pollinators, and feral honey bees in Argentine'Chaco Serrano'. *Ecological applications*, 4(2), 378-392.
- Arena, M., & Sgolastra, F. (2014). A meta-analysis comparing the sensitivity of bees to pesticides. *Ecotoxicology*, 23, 324-334.
- Greenleaf, S. S., Williams, N. M., Winfree, R., & Kremen, C. (2007). Bee foraging ranges and their relationship to body size. *Oecologia*, 153, 589-596.
- Corona, M., Libbrecht, R., & Wheeler, D. E. (2016). Molecular mechanisms of phenotypic plasticity in social insects. *Current opinion in insect science*, 13, 55-60.
- Meeus, I., Brown, M. J., De Graaf, D. C., & Smagghe, G. U. Y. (2011). Effects of invasive parasites on bumble bee declines. *Conservation Biology*, 25(4), 662-671.
- Soroye, P., Newbold, T., & Kerr, J. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science*, 367(6478), 685-688.

Data

- Seltmann KC, Allen J, Brown BV, Carper A, Engel MS, Franz N, Gilbert E, Grinter C, Gonzalez VH, Horsley P, Lee S, Maier C, Miko I, Morris P, Oboyski P, Pierce NE, Poelen J, Scott VL, Smith M, Talamas EJ, Tsutsui ND, Tucker E (2021) Announcing Big-Bee: An initiative to promote understanding of bees through image and trait digitization. *Biodiversity Information Science and Standards* 5: e74037. <https://doi.org/10.3897/biss.5.74037>
- California Department of Pesticide Regulation. (2017). Pesticide use reporting (PUR). California Environmental Protection Agency. Retrieved from <https://www.cdpr.ca.gov/docs/pur/purmain.htm>
- U.S. Census Bureau (2024). County Population Totals and Components of Change: 2020-2023 Retrieved from <https://www.census.gov/data/datasets/time-series/demo/popest/2020s-counties-total.html>
- Hoekstra, J. M., Molnar, J. L., Jennings, M., Revenga, C., Spalding, M. D., Boucher, T. M., Robertson, J. C., Heibel, T. J., & Ellison, K. (2010). *The atlas of global conservation: Changes, challenges, and opportunities to make a difference*. J. L. Molnar (Ed.). Berkeley: University of California Press.
- Earth Resources Observation and Science Center (2024). Annual National Land Cover Database. U.S. Geological Survey. Retrieved from <https://www.usgs.gov/centers/eros/science/data-access>