Investigating Lygus Spread in Driscoll's Raspberry Fields in Baja, Mexico

By: Joshua DeSouza, Kelsey Hope, & Bryce Stouffer

Professor Nicholas Cuba

ID 296: Advanced Vector GIS

May 3, 2019

Abstract

Serving as a final project for Advanced Vector, we are collaborating with Driscoll's and conducting research to aid the berry company's efforts in terms of investigating Lygus infestation of raspberry fields. Collectively, we answer the research question delivered from Driscoll's; does the proximity of strawberry fields to raspberry fields indicate the presence of Lygus in raspberry fields? We suspect that the more strawberry acreage around our raspberry fields, the more Lygus will be present in raspberry fields. Data from Driscoll's and satellite imagery from Landsat 8 enables us to evaluate our research objective through statistical modeling. The initial multiple linear regression showed no relation between strawberry acreage proximity to raspberry fields and lygus presence in raspberry fields. However, a logistic regression had more success predicting lygus presence based off strawberry acreage proximity to raspberry fields. The logistic regression could be improved with a larger sample of raspberry field spray data.

Problem statement/objectives

With a combination of entomology and spatial analysis, this study aims to answer the following research question:

Does the proximity of Driscoll's raspberry fields to strawberry fields affect the behavior of Lygus movement from one location to the other?

To help guide this question, the team will collaborate with Driscoll's, a California based seller of fresh strawberries and other berries. In collaboration with Driscoll's GIS Planning Analyst, Ahna Miller, together we aim to answer the problem statement that involves plant eating insect called Lygus. Lygus is a plant eating insect that has a diet of strawberries. Miller shares that there is an opportunity to understand the pest behavior and movement of Lygus based on the proximity of raspberry fields to strawberry fields. To take this opportunity, the team will aim to not only provide answers to the research question but additionally, gain insight on pest behavior of Lygus in the location of Driscoll's growing fields in Oxnard, California within a certain time period.

With GIS, one of the objectives is to provide a step closer to understanding the pest movement of Lygus in two locations. The study locations are Baja, Mexico and Oxnard, California. Another objective is to dive into the entomology of Lygus and understand the accounts of Lygus movement in both study areas. The main objective is to test whether the proximity of raspberry fields to strawberry fields is amplifying or inhibiting pest behavior. Ideally, the team would like to expect a correlation between the lygus presence and varying acreage of strawberry fields around raspberry fields.

With GIS and satellite data, this study will be able to select a certain time frame and create distance buffers around focal raspberry fields with strawberry fields in the surrounding farm-scape. These distance buffers will allow us to conclude, what distances amplify pest movement and then proceed to map out the farm-scape with buffers. With the Landsat 8 data, remote sensing will be needed to false color composite to identify local competitors and the location of their strawberry fields. Then the refined images can be used to help identify land cover types and perform GIS analysis. The outcome of this study is expected to use GIS for contributing to Driscoll's efforts in providing the freshest berries to the consumer world. With our research, GIS will help advance and catalyst new methods of pest control and provide a stepping stone to GIS planning at Driscoll's.

Data

SOURCE	USE	TYPE
Landsat 8	Competitor Strawberry Fields	.rst (raster)
Driscolls	Raspberry & Strawberry locations	.shp (vector)
Driscolls	Spray Data	Tabular
Driscolls	Defines Study Area	.shp (vector)

Methodology

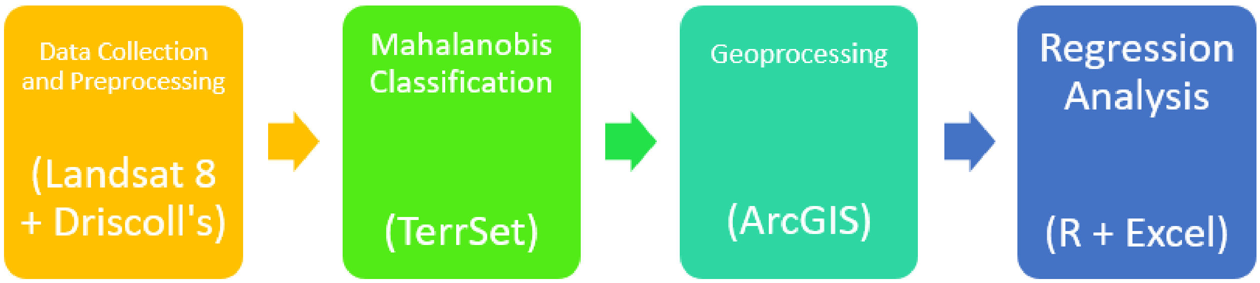
First, we collected data from Driscoll's and images from Landsat 8 within the specified time frame (June-September 2018). Then we displayed the images with a false color composite using bands 4, 5, and 7 (R, G, B). Next, we will use image classification to classify Driscoll competitors' strawberry fields.

The classification technique in use is Mahalanobis Soft Classification. The Mahalanobis classifier calculates similarity between the spectral characteristic of a pixel and the spectral characteristic of the training sites, using Mahalanobis distances. Unique to other classifiers, this technique is useful when classifying a single class. Furthermore, Mahalanobis distances range from zero to infinite, and are then transformed into a measure of typicality, ranging from 0 to 1. Our research team considers 0.15 the threshold of typicality when classifying strawberry fields.

The classification output identifies all strawberry fields within the study area; however, we specifically need a shapefile of competitor strawberry fields. Therefore, using a select by location we can accomplish creating a shapefile of competitor strawberry fields.

Second, we use a geoprocessing tool, Multi Ring Buffer Analysis, to set several buffers (150m, 1,000m, 1,500m) around the raspberry fields, as instructed by Driscoll's. Then the team joins the spray tabular data based on raspberry field ID to determine which raspberry fields have probable Lygus presence. The last geoprocessing steps involve a series of dissolves and intersects to prepare independent and dependent variables for the regressions. After processing the variables to all be in the same table in ArcMap, the team organizes the variables in Excel to streamline the regressions. The independent variables consist of strawberry acreage within each buffer zone, and the dependent variable is based off probable spray data (*Figure 4*).

Next, the team utilizes two regressions: multiple linear regression, and logistic regression. The team uses Real Statistics Resource Pack to carry out the multiple linear regression in Excel. Then, we choose to run a logistic regression in addition to a multiple linear regression because we have very few variables and their distributions were not normal. In addition, logistic regressions model binary outputs well, and determining the presence or absence of Lygus in raspberry fields is a binary venture. In order to operate the logistic regression analysis, we reclassed the spray data (probable sprays). If there were more than zero sprays, we allocate a value of one (Lygus presence). Conversely if there are no sprays, the value remains zero (Lygus absence). The team uses the “glm” function from the stats package in R to run the logistic regression.



Results

The multiple linear regression indicated no relationship between the amount of strawberry acreage in proximity to raspberry fields and presence of Lygus in Driscoll’s raspberry fields. The Adjusted R Square value seen in *Figure 4* is -0.0756 for the multiple linear regression. *Figure 4* shows the 150-meter strawberry buffer acreage to be the strongest independent variable of the three with a coefficient of 0.004955. However, its p-value is 0.476, which proves its power as a coefficient in the model is statistically insignificant.

The logistic regression proves to be a better way of modeling the relationship between the amount of strawberry acreage in proximity to raspberry fields and presence of Lygus in raspberry fields. This model uses the binary Lygus presence data, which is a conversion from the probable spray count data. The logistic model gives one field a low probability of Lygus presence when the field has Lygus based on our binary conversion of “Probable Spray Data” (*Figure 6)*. It does well to predict the probability of the final nine fields with Lygus spray data. The strawberry acreage within 1500 meters of raspberry fields proves to be the strongest coefficient in this model and is close to being statistically significant with a p-value of .0992 (*Figure 7)*. However, the model shows lack of capability due to it totally missing the first field and due to it giving mid-range probabilities of Lygus Presence. More samples of field spray data could be a way to improve this model. Nevertheless, the logistic regression is a better way to model Lygus presence due its binary nature.

Conclusion

The hypothesis expected that more strawberry acreage around the raspberry fields, the more Lygus presence in raspberries. However, the research strategy (*finding correlation between lygus movement and the frequencies of insecticide sprays*) was ineffective and inconclusive. The correlation analysis did not find any relationship between the size of strawberry fields to the presence levels of lygus in raspberry fields.

The study was able to accomplish some outputs. This research classified strawberry fields in Baja, Mexico to identify the ratio of competitor fields to Driscoll's strawberry fields. The Mahalanobis soft classification allowed us to identify strawberry fields to geocode Lygus presence levels. The study used Multi Ring Buffer analysis to identify whether the proximity of raspberry to strawberry fields contributes to lygus presence. Initially, multivariate regression analysis indicated no correlation between the area and proximity of strawberry fields with lygus presence. After finding no correlation, the team tried logistical regression to test out whether the area of strawberry fields within distance buffers can accurately predict the presence of lygus. The results detailed to be a good model fit, and there is probable explanation of dependent variables by the strawberry field size (acreage). Overall, the study was unable to answer the research question and hypothesis with our research strategy given by Driscoll's. This study should consider alternate methods or research questions to gain a better understanding of Lygus presence in raspberry fields in Baja Mexico.

The study's results could be due to the limited data provided by Driscoll's. The multivariate regression analysis did not contain enough explanatory variables to model the relationship with Lygus presence levels. The original spray data required pre-processing for the regression analysis for the missing values or "null values" that needed input. To fix that, the team assumed that null or missing values can be replaced with a value of zero, to represent "zero lygus presence". At varied distances, the regression analysis could have modeled the relationship better if more spray data was given to the team. In addition to the data, the classification of strawberry fields could have affected the results. A big question is what is the accuracy of the classified strawberry fields. The Mahalanobis classifier was reasoned to be the best method to identify the spectral characteristics of the strawberry fields. The research considered 0.15 for the threshold of typicality because this value will be inclusive when identifying the strawberry fields. If the threshold of typicality was larger, the study risks the exclusion of strawberry fields in the classification output. The accuracy of the classification was not tested nor repeated to test out different thresholds of typicalities. The study's results could be affected by the threshold value of 0.15 and the study would have excluded strawberry fields owned by competing companies. The study will need to definitely consider, for further research, the testing of different thresholds for classifying strawberry fields.

Despite these setbacks, the study has the potential to explore and conduct deeper investigation of pest behavior and movement on the Lygus bug. To understand Lygus, the study could have benefitted from extensive literature review on Lygus and exploring methods to biologically control the lygus from entering raspberry fields. The study will need to improve or create a new research strategy to modify this study in Baja, Mexico. The study could consider further research into Lygus. Although, this study made references to a few studies that were situated in California, no references were made to research of lygus movement in Baja Mexico. In addition to the no contextual reference of lygus movement in Baja, the study will need to invest in gathering more spray data. This study recommends that Driscoll's invests in data collection of Lygus spray data for further research and to modify this study.

References

Dara, S. (n.d.). *Lygus bug and its management in strawberries.* Santa Maria: University of California Cooperative Extension.

Goodell, P. B. (2001). *Managing Lygus in the Landscape.* Fresno: California Chapter A, Agronomy Society.

Swezey, S. L., Nieto, D. J., Hagler, J. R., Picket, C. H., Bryer, J. A., & Machtley, S. A. (2013). Dispersion, Distribution, and Movement of Lygus spp. (Hemiptera: Miridae) in Trap-Cropped Organic Strawberries. *Environmental Entomology*, 770-778.

Figures

Figure 1: Study Area Map for Lygus Project in Baja California, Mexico

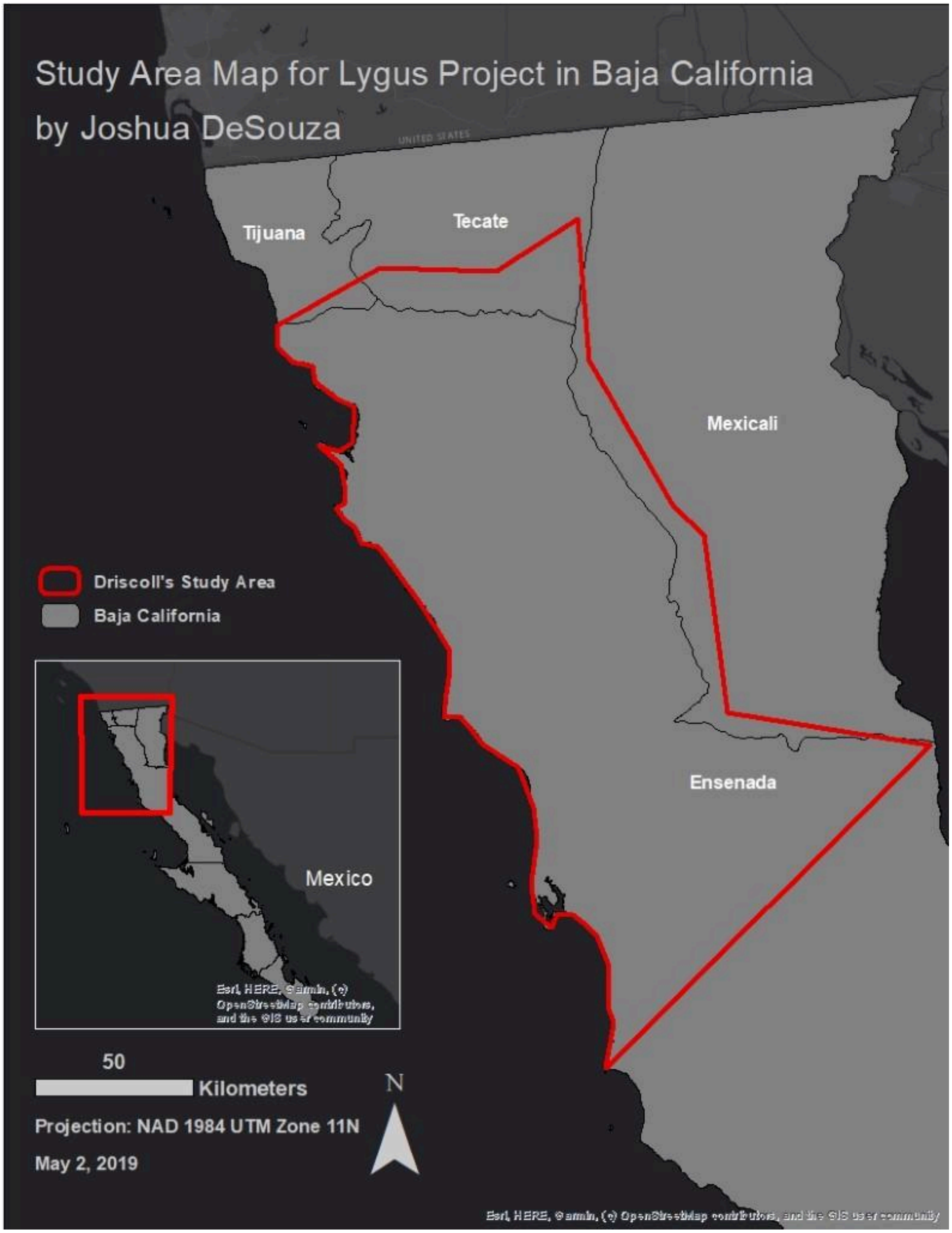


Figure 2: A map indicating Lygus Presence in raspberry fields based on spraying data

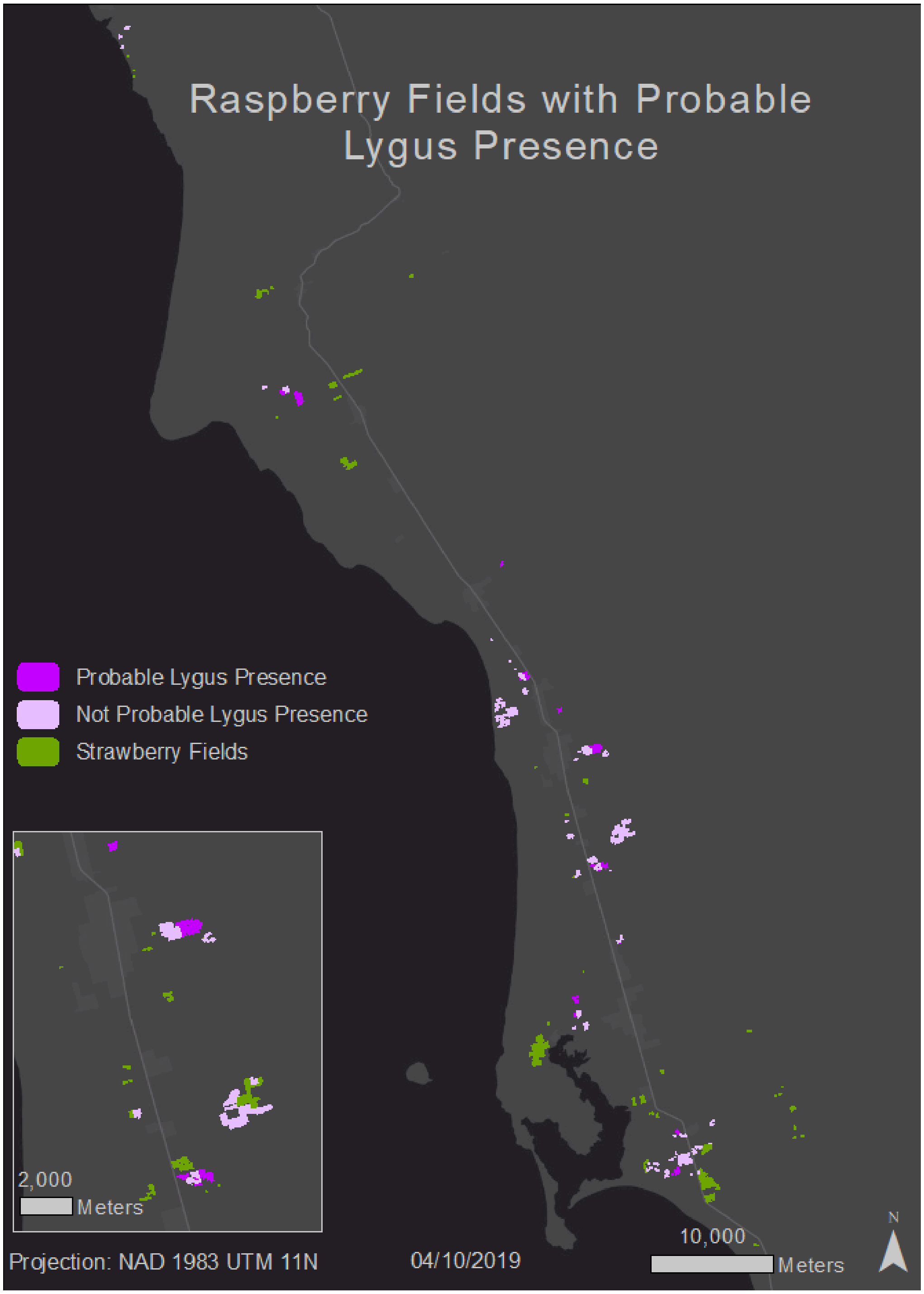


Figure 3: Buffer Zones Around Raspberry Fields at Three Increments

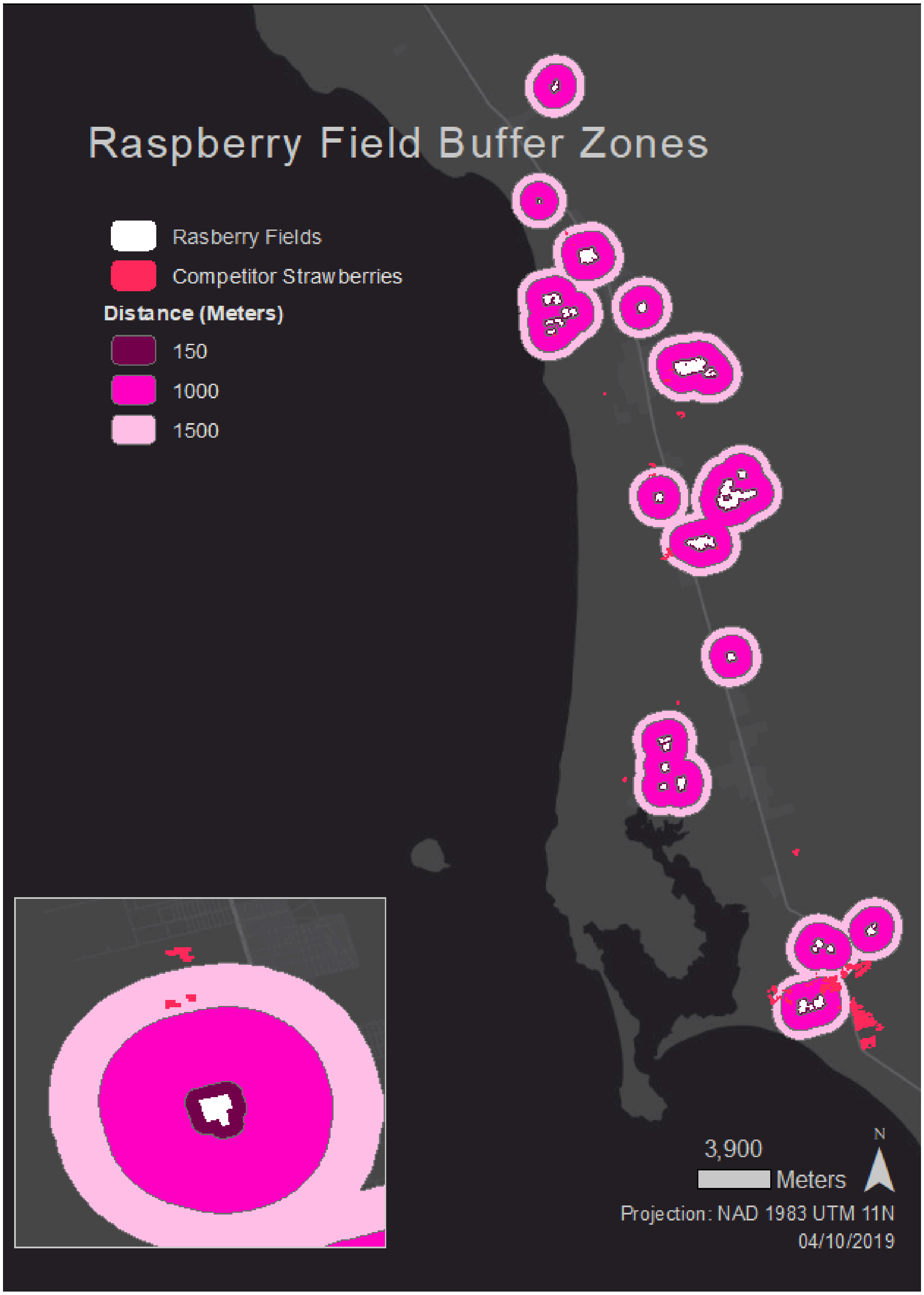


Figure 4: Variables used in Regressions

Field ID	150m Buffer Acreage	1000m Buffer Acreage	1500m Buffer Acreage	Presence Based on Spray Data	Field Spray Count
10332	0.00	2163.17	2712.22	1	7
9070	0.00	890.97	1247.15	1	6
10197	395.47	1247.15	1247.15	1	5
10208	0.00	2149.20	2190.17	1	5
10194	0.00	809.58	1263.73	1	4
10822	0.00	73.88	228.35	1	4
10823	0.00	82.40	226.94	1	4
10175	0.00	54.01	54.01	1	3
10824	0.00	127.91	233.94	1	3
9083	14.00	2394.02	4062.85	1	1
10177	44.99	118.93	604.98	1	1
9068	0.00	0.00	54.01	0	0
10246	0.00	300.75	300.75	0	0
10331	0.00	454.77	454.77	0	0
10333	0.00	2151.42	2190.17	0	0
10806	7.78	300.75	308.64	0	0
10807	0.00	300.75	300.75	0	0

Figure 5: Multiple Linear Regression Statistics

Regression Statistics	
Multiple R	0.355083721
R Square	0.126084449
Adjusted R Square	-0.075588371
Standard Error	2.516922556
Observations	17

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1.809204276	0.85338643	2.120029	0.053818
150m Buffer Acreage	0.004955088	0.006757584	0.733263	0.476414
1000m Buffer Acreage	0.000995462	0.002332181	0.426837	0.676482
1500m Buffer Acreage	-0.000203882	0.001781458	-0.11445	0.910633

Figure 6: Logistic Regression Plot

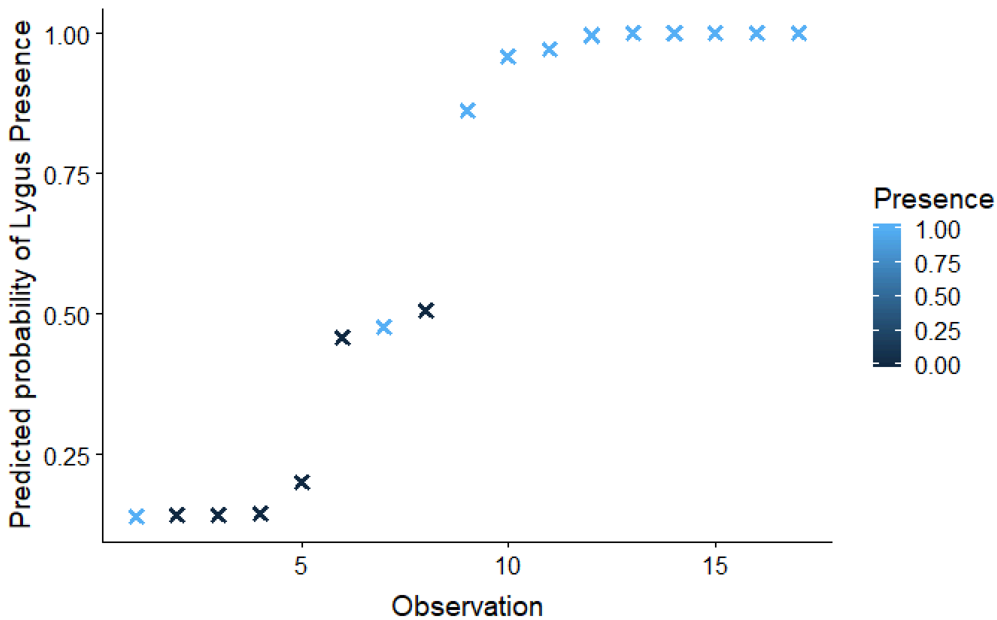


Figure 7: Logistic Regression Statistics

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.83579    1.30755  -1.404   0.1603
`150`        0.01813    0.04021   0.451   0.6522
`1000`      -0.03423    0.02086  -1.641   0.1009
`1500`       0.03439    0.02086   1.649   0.0992
```

Appendix

Literature Review

Given that Lygus is such a common pest to many cash crops (alfalfa, beans, cotton, strawberries, apples, pears, and other cool-season vegetables), there has been a fair amount of interest in researching methods for reducing the insects' impact, and its movement throughout agricultural fields. Perhaps most important among the findings as it relates to our project is that Lygus moves around from plants and crops quite often; as soon as its habitat becomes unsuitable (possibly from complete consumption of available nutrition), it moves to more suitable areas (Goodell, 2001). Given that this knowledge, many researchers have investigated Lygus movement prevention or disruption. They have found that habitat diversification through interplanting in fields reduces movement. It is recommended to preserve some Lygus habitat by retaining uncut or unharvested strips of crops; these areas can act as a buffer for other parts of the crop fields. This is because "Lygus infestations have been found to be more severe when fruit is continuously present in the agricultural fields" (Dara). Alfalfa is a recommended crop to use as a buffer, as it is highly attractive to Lygus (Swezey, et al., 2013). Research has found that Lygus were 8 times more abundant in trap crops compared to interior strawberry rows. Furthermore, "when Lygus movement out of trap crops does occur, it is mostly limited to the strawberry rows in closest proximity to the alfalfa" (Swezeu, et al., 2013). Knowing this, and with additional research we can help Driscoll's determine proper buffer distances between berry fields to reduce the movement of Lygus pests.

Joshua's Reflection:

For this course, I enjoyed connecting the geo-processing techniques such as multi-ring buffering, and spatial statistics for the Lygus Strawberry project. The project has taught me a couple of things. As a multi-disciplined project, I really enjoyed stringing together the different techniques from remote sensing and from advanced vector GIS to answer the research question. The sequence of steps from geo-processing to conducting logistic regression required my ability to understand where to apply different data inputs. With many data sources, the practice of data management and processing was another thing I enjoyed. The art of organizing data for further processing or analysis was rewarding for me. With previous experience with excel, I was able to organize the Spray data for logistic regression and conduct the analysis smoothly. With Professor Cuba's support, the explanatory variables were identified which helped my organization of data. This project overall was very rewarding and I was happy to work with Bryce and Kelsey for this project. Everyone was able to apply their unique strengths in this project and I was grateful for the turnout.

After observing the other presentations, the following considerations were inspired for further research in this project. Firstly, I noticed that all of the presentations had exploration of other types of regression analysis. OLS Regression Analysis was a popular tool used for the impacts of solar panels, the national borders, and other phenomena. The presentations inspired me to consider other GIS processing methods and techniques. The study did not explore the biology of Lygus and that migration behaviors. The study also did not consider the research into the classification methods for strawberry fields. The team and I realized that extending the research to figure out whether climate change is affecting the movement of lygus. The presentation also inspired the team and I to commit to research on Baja, Mexico and whether climate change is affecting the movement of lygus. The presentations also inspired the team and I to review and modify the study for better results and to test the accuracy of the research.