Predicting plant-pollinator interactions using flower abundance

Marissa Childs

August 22, 2015

Abstract

Logistic regressions were fit to existing flower abundance and plant-pollinator interaction data sets collected in 2011-2014 in the H.J. Andrews Experimental Forest. The model fit with four flower species, *Gilia capitata*, *Eriophyllum Lanatum*, *Hypericum perforatum*, and *Ligusticum grayi*, and no hierarchical structure found the probability of interactions between *Apis mellifera* and the target species increased with increasing abundance of the target species and there was no statistically significant negative effect on the probability of an interaction caused by the presence and abundance of other flower species. For one flower species, *Gilia capitata*, the abundance of another flower species, *Eriophyllum lanatum*, was found to increase the probability of an *Apis-Gilia* interaction. On the other hand, a logistic regression fit with fifteen flower species that incorporated the hierarchical structure of the data found that on the level of the meadow network, the probability of interactions between *Apis mellifera* and the target species increased with increasing abundance of the target species for only a three target species, none of which were among four flower species in the initial model. On the other hand, the hierarchical model found many more statistically significant negative effects of competitor plant species and no positive statistically significant effects. An attempt to rank the relative preference *Apis* has for various flower species was inconclusive, using both the regressions from the four species model and the 15 species model.

Introduction

Plant-pollinator networks are highly important ecological systems whose structure and resilience to disturbances is still not entirely understood. Montane meadows support a very high diversity of flowering plants and pollinator species and thus the loss of meadow area is a cause for some concern. The area of the montane meadows in the H.J. Andrews Experimental Forest has been shrinking since 1948 (Highland, S. 2011), so an effort has been made to better understand these meadows. Data has been collected on the plant-pollinator networks in the montane meadows of the H.J. Andrews Experimental Forest in Oregon by the Ecosystem Informatics Summer Institute from 2011 to 2015. In each meadow, both flower abundance and plant-pollinator interaction data was collected in ten 3 meter by 3 meter plots. For each year a meadow was surveyed, five to seven 15-minute watches were carried out for each of the ten plots in the meadow. A single watch consists of a survey of all plant species in the plot with flowers in anthesis and the abundances of plant species that are in anthesis and 15 minutes of observation of the interactions between plants and pollinators in the plot. In 2011, fifteen meadows were included in the data collection but 6 of those meadows were outside boundary of the H.J. Andrews Experimental Forest so beginning in 2013, those 6 meadows were not visited and instead 3 new meadows were added. In the first 4 years of data collection, nearly 500 pollinator species and 130 plant species were identified.

In the H.J. Andrews experimental, *Apis mellifera* has been present for an unknown length of time, yet it is non-native so it was certainly introduced to the area, possibly when honeybees were brought in to pollinate nearby crops. The effects of *Apis mellifera* on native pollinators in the H.J. Andrews is still relatively unknown. Other studies suggest that non-native species can disrupt plant reproductive mutualisms and, in particular, noted that *Apis* have, in multiple regions, had significant effects on the native bees involved in plant-pollinator mutualisms (Traveset, A. and Richardson, D.M. 2006). None of the regions considered by Traveset and Richardson are near the H.J Andrews so the effects of *Apis mellifera*, both presently and in the long-term, are still unknown. One route to better understanding the effect of *Apis mellifera* is to consider the pref-
ferences of *Apis mellifera*, especially considering how those preferences interact with those of native pollinator species. *Apis mellifera* has been shown to have som preferences despite being a generalist: a study in a controlled indoor environment showed that *Apis mellifera* is preferred small pollen grains and that odor was the main cue used by bees to select pollen (Pernal, S.F. and Currie, R.W. 2002).

In this paper I attempted to both discover a preference ranking for the flower species found in the H.J. Andrews meadows and also understanding what variables can explain the occurrence of specific plant-pollinator interactions. To do so, I looked at the interactions of *Apis* with various flower species and the abundances of those flower species when interactions did or did not occur. Such a model would give us the ability to predict whether or not a pollinator species will interact in a meadow. The choice of *Apis* was twofold: first, *Apis mellifera* is a generalist pollinator species that, within the data set, has interacted with 50 different plant species, providing significantly more data for analysis. Second, *Apis mellifera* is preferred small pollen grains and that odor was the main cue used by bees to select pollen (van Engelsdorp, D. and Meixner, M.D., 2010). Therefore, increased knowledge of *Apis mellifera* preferences can help inform both management of meadow networks and management of industrial honeybee colonies.

This paper attempts to investigate the following questions:

1. Is the occurrence of a visit between *Apis mellifera* and a specific plant species in a meadow a function of the abundances of that plant species?
2. Does the presence of other plant species, i.e. competition, have an effect on the occurrence of a visit between *Apis mellifera* and a target plant species?
3. Does *Apis mellifera* have a preference between various plant species and how can the preference *Apis mellifera* has for a set of plant species be quantified or ranked?

**Methods**

Logistic regression models were fitted to data on plant-pollinator interactions first for four plant and *Apis mellifera*, and later, for 15 plant species and *Apis*. Visits of *Apis mellifera* to a target plant species was the dependent (0,1) variable. The model was fit twice with maximum likelihood estimation: in one fit of the model the independent variable was only the abundance of the target species and in another fit of the model, the independent variables were the abundances of flowers of all four plant species. The model was also fit with the hierarchical structure of the data incorporated to the statistical analysis.

For multiple reasons the analysis in this paper focuses on interactions at the meadow level. The meadow data tended to be less variable, or noisy, because it was aggregated over all of the plots observed in that meadow. Furthermore, meadows were considered the ecologically significant spatial unit. Thus, each data point is a meadow in a specific watch in a year, or a meadow-watch-year (MWY).

To establish some notation: let $Y_{i,j,m,w,y} \in \{0,1\}$ be 1 if pollinator $i$ interacts with plant $j$ in meadow $m$ during watch $w$ of year $y$ and 0 otherwise. Additionally, let $A_{j,m,w,y}$ be the aggregated abundance of plant $j$ in meadow $m$ during watch $w$ of year $y$.

**Model fitting with frequentist generalized linear model**

I fit two different logistic regressions for each of four target flower species in R using maximum likelihood estimation. Only four flower species were used in this model because it was intended to be a test of how well the model work. I intentionally chose the plant species to be ones with which *Apis* has been observed to interact at varying frequencies. When ranking flower species by the number of plots in which *Apis mellifera* interacted with them, *Gilia capitata* was first, *Eriophyllum lanatum* second, *Ligusticum gray* sixth, and *Hypericum perforatum* seventh.

In the first model, for target plant species $j^*$, I fit

$$
\log\left(\frac{P(Y_{i,j^*,m,w,y} = 1)}{1 - P(Y_{i,j^*,m,w,y} = 1)}\right) = \beta_0 + \beta_{j^*} A_{j^*,m,w,y}.
$$

That is, for each of the four flower species, I fit a model where the only explanatory variable was the abundance of the target flower species.

For the second model, for each plant species $j^*$, I used

$$
\log\left(\frac{P(Y_{i,j^*,m,w,y} = 1)}{1 - P(Y_{i,j^*,m,w,y} = 1)}\right) = \beta_0 + \sum_{j \in J} \beta_j A_{j,m,w,y}.
$$
This curve incorporated the flower abundances of the other three plant species as explanatory variables in addition to the abundance of the target species.

For both model fits for each target plant species, I calculated McFadden’s pseudo $R^2$ (McFadden, 1974) by extracting the log of the maximum estimated likelihood for the intercept model and full model and using the formula

$$R^2 = 1 - \frac{\log(\hat{L}(M_{Full}))}{\log(\hat{L}(M_{Intercept}))}$$

where $\hat{L}$ is the maximum estimated likelihood.

To gain a measure of the relative preference *Apis mellifera* had for each plant species, I used the model fit with all four flower abundances for each target species to calculate the derivative with respect to the target abundance at $P(Y_{i,j,p,m,w,y} = 1) = 0.5$, the abundance of the target species at which $P(Y_{i,j,p,m,w,y} = 1) = 0.5$, the $y$-intercept, or the probability when the abundance of the target species was zero, $P(Y_{i,j,p,m,w,y} = 1)$ at the maximum observed abundance of the target species, and $P(Y_{i,j,p,m,w,y} = 1)$ at the mean observed abundance of the target species.

**Model fitting with Bayesian hierarchical model**

When I used a model that incorporates the hierarchical nature of the data collection, I also expanded my analysis to all plant species that *Apis mellifera* was observed interacting with in at least 10 plots over the first three years of data collection in the nine meadows that were surveyed for all years. The flowers included was expanded because, after running the first model, few effects of competition were found and I felt one possible explanation of this was that the flower species chosen did display the effects of competition between them but other flower species might. At this point, I held out all data collected in 2014 from the model fitting process so as to reserve some data to test the fitted model on. I also limited the data to only meadows that were observed for all years, as I was interested in consistent data across years and some meadows were only observed in 2011-2012 while others were only observed from 2013 on.

The hierarchical model is defined in the following way: let $\bar{a}_{m,w,y}$ be a vector of the flower abundances for all of the plant species under consideration in meadow $m$ in watch $w$ in year $y$ and $\bar{\beta}_m$ be a vector of the coefficients of the abundances for meadow $m$. Then for a target plant species $j$,

$$\log \left( \frac{P(Y_{i,j,m,w,y} = 1)}{1 - P(Y_{i,j,m,w,y} = 1)} \right) = \bar{\beta}_m \cdot \bar{a}_{m,w,y} + \beta_0,$$

where

$$(\forall m)\bar{\beta}_m \sim \text{Normal}(\mu_{\vec{\beta}}, \sigma_m),$$

$\vec{\sigma} \sim \text{Uniform}(0, 10),$

$\mu_{\vec{\beta}} \sim \text{Normal}(0, \mu_{\sigma}),$ and

$\mu_{\sigma} \sim \text{Uniform}(0, 10)$

with $\sigma_m$ the element in $\vec{\sigma}$ that corresponds to meadow $m$.

This hierarchical model is based on the premise that, for a single target flower species, in each meadow, across all years and watches, the coefficients of the abundances are similar. In turn, we expect all of the meadow level betas, the $\beta_{m,s}$, so be similar, and thus expect them all to be drawn from a normal distribution with a shared mean, $\mu_{\vec{\beta}}$, which is the coefficient at the level of the entire H.J. Andrews Forest. Since $\mu_{\vec{\beta}}$, $\vec{\sigma}$ and $\mu_{\sigma}$ are used to describe the distributions of other variables, they must each either be specified or given distributions; I specify uninformative prior distributions for each. As the variation between meadows level $\beta_{m,s}$ and between $\mu_{\vec{\beta}}$ was not large, uniform distributions allowed me to give uninformative priors with small ranges for $\vec{\sigma}$ and $\mu_{\sigma}$, their respective standard deviations, but still give no preference as to where the variables would fall within the 0-10 range. For the distribution of the $\mu_{\vec{\beta}}$, since each element of the vector could have positive or negative values and I had no prior knowledge as to what that value was, a normal distribution with a mean of zero was an uninformative prior was used.

Using RStan, a probabilistic programming language for R, I fit this model to the data collected in 2011-2013. There were 15 flower species that were visited in over 10 plots by *Apis mellifera* in the three years of interaction data: *Gilia capitata*, *Eriophyllum lanatum*, *Erigeron foliosus*, *Potentilla gracilis*, *Ligusticum grayi*, *Hypericum perforatum*, *Agoseris heterophylla*, *Eriogonum compositum*, *Sedum oreganum*, *Solidago canadensis*, *Agoseris heterophylla*, *Eriogonum umbellatum*, *Rudbeckia occidentalis*, *Achillea millefolium*, *Rumex acetosella*. For each target species, I fit the above model with a the explanatory variables of abundance of all 15 flower species as well as a null model with only an intercept, running for 100,000 iterations with 4 chains, thinning by 50.
\( \hat{r} \) values were considered as a measure of convergence.

As a measure for goodness of fit, I used a variation on McFadden's pseudo \( R^2 \). I used the mean of the log posterior for the full model and the mean of the log posterior for the null model as proxies for the log of the maximum likelihood of the full model and the log of the maximum likelihood of the null model, respectively.

To evaluate the predictive power of the model, I looked at the heldout data from 2014 and calculated the probability of an Apis interaction with each of the target flower species in each of the meadow watches using the flower abundance data and the estimated meadow level coefficients. I then plotted the precision-recall curve, where recall is the portion of observed interactions that occurred that were correctly predicted by the model and precision is the portion of predicted interactions that were observed to occur. Each point one the curve is calculated by considering various threshold probabilities.

**Results**

For *Gilia capitata, Eriogonum lanatum, Hypericum perforatum* and *Ligusticum grayi*, the probability of an Apis mellifera-interaction occurring increased with abundance of the flowers with \( p < 0.0001 \) for all four target species and pseudo \( R^2 \) values of 0.5764, 0.2823, 0.2301, and 0.2238 for each of the target species, respectively. (Figures 1a, 1b, 2a, and 2b).

At the meadow level, *Apis-Gilia* interactions increase with abundance of *Gilia* \( p < 0.0001 \), and also increase when *Eriogonum lanatum* is present \( p < 0.004 \) (Table 1). At the meadow level, *Apis-Hypericum* interactions increase with abundance of *Hypericum* \( p < 0.0001 \), and are not related to the presence of other species (Table 1). At the meadow level, *Apis-Eriogonum* interactions increase with abundance of *Eriogonum* \( p < 0.0001 \), and also increases when *Gilia* is present \( p < 0.06 \) (Table 1). At the meadow level, *Apis-Ligusticum* interactions increase with abundance of *Ligusticum* \( p < 0.0001 \), and are not related to the presence of other species (Table 1).

The threshold at which interactions of *Apis* with flowers increased was lowest, the slope of the relationship was highest, the probability at the maximum observed abundance was highest, and the probability at the mean observed abundances was highest for *Gilia capitata* (Table 2).

In the fits of the hierarchical model for each target flower species, the highest \( R \) observed was 4.4 but most fits had \( R_s \) between 1 and 1.5. The pseudo \( R^2 \) values ranged from as high as 0.9674 for the model fit for *Eriogonum compositum* to as low as 0.0950 for *Achillea millefolium* and had a median value of 0.2152119. The only \( \mu_3 \) that was positive and statistically seperated from zero occured for *Eriogonum foliosus, Angelica arguta,* and *Agoseris heterophylla* when that flower species was the target species (Figure 3). All other \( \mu_3 \) coefficients were not statistically seperated from zero or negative and statistically seperated from zero (Figure 3). For a target species of *Gilia capitata*, the meadow level coefficients on *Gilia capitata* abundance are positive or not seperated from zero and the meadow level coefficients on other flower speices are negative or not seperated from zero, except in the case of the coefficient for *Potentilla gracis* in the M2 meadow (Figures 4). For a target species of *Agoseris heterophylla*, the meadow level coefficients on *Agoseris heterophylla* abundance are positive or not seperated from zero, and the meadow level coefficients on other flower speices are negative or not seperated from zero, except in the cases of the coefficient for *Eriogonum numbellatum* in meadow RP1 and *Eriophyllum lanatum* in RP1 and RP2 (Figure 5).

Despite the positive significant coefficients seen on the target flower species abundance with the 4 species non-hierarchical model, none of the four flower species had a positive significant \( \mu_3 \) coefficient in the 15 species hierarchical model. The hierarchical model also found that there was no significant effect of the abundances of other flower species between *Gilia capitata, Eriogonum lanatum, Hypericum perforatum* and *Ligusticum grayi*. In fact, the model fits for *Eriophyllum lanatum, Hypericum perforatum* and *Ligusticum grayi* as target species were all found to have no statistically significant \( \mu_3 \) coefficients and the model for *Gilia capitata* as target species had only a significant coefficient on *Achillea millefolium* which was negative.

The model fit with the hierarchical model has a precision of approximately 50\% regardless of the threshold used but to reach recall levels above 15\%, thresholds had to be very close to zero. The area under the precision-recall curve was 0.5121 (Figure 6).

---

*Predicting plant-pollinator interactions*  
Childs, M.
Discussion

For each plot, the total number of flowers of each species is calculated using the total number of stalks in the plot multiplied by the average number of flowers per stalk in that plot. Analysis of the flower abundance data has shown that the number of stalks and flowers per stalk often vary significantly between years and observers, more so than might be expected from natural variation. However the distribution of total abundance, while still varying by observer and year, does not display as severe fluctuations (Sander et al., 2015). This is a cause for concern as the models in this paper rely solely on flower abundance data as explanatory variables for interactions.

When fitting the model I also realized that for a plant species $j^*$, if the actual abundance of flowers of species $j^*$ in a plot is $A_{j^*,p,m,w,y} = 0$, then $P(Y_{i,j^*,p,m,w,y} = 1) = 0$ since pollinator $i$ cannot land on a flower of species $j^*$ if the flower does not occur in the plot. Yet with the log odds model, $P(Y_{i,j^*,p,m,w,y} = 1) = 0$ only when
\[
\frac{\beta_0 + \sum_{j \in J} \beta_j A_{j,p,m,w,y}}{1 + e^{\beta_0 + \sum_{j \in J} \beta_j A_{j,p,m,w,y}}} = 0,
\]
that is, only when $\beta_0 + \sum_{j \in J} \beta_j A_{j,p,m,w,y} = -\infty$.

Mathematically, model will never predict such an occurrence, so with the logistic regression, $P(Y_{i,j^*,p,m,w,y} = 1) = 0$ will never occur. However $P(Y_{i,j^*,p,m,w,y} = 1) = 0$ can numerically occur in R due to rounding. For all four of the flowers initially used to fit the model, when the total abundance of the target flower was zero, the probability of an interaction was $P(Y_{i,j^*,p,m,w,y} = 1) < 0.1$.

Another problem encountered while fitting a hierarchical model was that the $n_{\text{effective}}$ was much lower than desired. A small $n_{\text{effective}}$ implies that the posterior draws are highly correlated. Independent draws are desired, so thinning can be used to help counteract this. When I ran RStan to fit the model for each target species, I ran with thinning every 50 but even this seems to have been insufficient. With more time or more computing power, each model fit could have been run for more iterations, thinning by a larger number. In the analysis from these fits, though, the low $n_{\text{effective}}$ means that estimates of standard error are low. This in turn will cause the calculated confidence intervals to be too narrow so coefficients may be found to be statistically separated from zero when they are not actually.

Effects of target flower species abundance

The model with no hierarchical structure suggests that there is a positive relationship between the abundance of the target flower species and interactions for all four flowers considered while the hierarchical suggest that there is a positive relationship between the abundance of the target flower species and interactions for some target flower species and no relationship for other target flower species.

For the four plants initially considered, the mean of the distribution of abundances for the flower species in question was higher in the MWYs where an interaction occurred than in the MWYs where no interaction occurred. For each of the four target flower species for both the model fit with only the target flower abundance as an explanatory variable and for the model fit with the abundances of all four flowers as explanatory variables, the target flower abundance coefficient was positive and statistically different from zero at the 5% level. However in the hierarchical models fit with 15 flower species abundances as the explanatory variables, I found only three $\mu_\beta$ coefficients of the target species that were statistically significant at the 5% level, none of which were the four flower species previously found to have statistically significant coefficients on own abundance.

There are many possibilities for why this occurred. First, the hierarchical model takes into account that each watch in a meadow is a repeated sample and therefore the repeated sample from a meadow would be expected to be more similar to each other than samples from different meadows. Yet I may have found no statistical significance of the coefficients of the target species because our model may not appropriate for the data and phenomenon being observed as, even at the meadow level, very few, if any, of the $\beta$ coefficients on own abundance for Gilia capitata, Erriphylum lanatum, Hypericum perforatum and Ligusticum grayi were statistically significant. Perhaps the hierarchical model should have looked at $\beta$ coefficients for meadow years as the intermediary step rather than meadows. Heldrop (2015) found that between years, there was significant turnover in the flower composition of a meadow which might support the use of meadow-years instead of meadows as the lowest level of $\beta$ coefficients. Second, the addition of other flower species may have caused the coefficients of the target flower’s abundance to no longer be statistically significant. If some of the added flowers are correlated in abundance to the
abundances of *Gilia capitata*, *Eriophyllum lanatum*, *Hypericum perforatum* or *Ligusticum grayi*, then a decrease in the coefficient, and therefore decrease statistical significance of the coefficient, would be expected.

**Effects of others flower species abundances**

When looking only at four plant species, I did not find a statistically significant negative effect of competition on the occurrence of an interaction between *Apis mellifera* and a given plant species on the 5% level. If the presence of plant species B had a negative affect on the probability of an *Apis mellifera* visit to plant species A, that is, when flowers from B were also present *Apis mellifera* was attracted away from plant species A, then I would expect a negative coefficient for plant species B in the fit of the model for plant species A. If flowers from B did not attract *Apis mellifera* away from A, I would expect a coefficient not statistically separated from zero. For the logistic regressions for *Ligusticum grayi* and *Hypericum perforatum* the only abundance coefficient significant at a 5% level was the abundance of the target plant species (Table 1). In the regression for the occurrence of an *Apis mellifera* visit to *Gilia capitata* the coefficient of *Eriophyllum lanatum* was positive and statistically different from 0 at a 5% level. While the coefficient for the abundance of *Gilia capitata* was similarly positive in the regression for *Eriophyllum lanatum* as the target species, the coefficient is not significant at the 5% level.

In the hierarchical model, many competing plant species abundances had statistically significant negative coefficients at both the meadow and meadow network levels, but the majority of the competing plant species abundances at the meadow network level had coefficients not statistically different from zero and at the meadow level, some coefficients for the abundances of competing flower species were positive at a statistically significant level. Both *Gilia capitata* and *Eriophyllum lanatum*, the coefficient of the other’s abundance at the both the meadow and meadow-complex level was not statistically significant.

The occurrence of positive coefficients on the abundances of competing species has multiple possible explanations. One is that the abundance of the competing species is correlated with the abundance of the target species and is therefore found to have a positive coefficient. Another possible explanation is that the presence of plant species B increases the likelihood of an interaction between plant species A and *Apis mellifera* because plant species B tends to occur when and where *Apis* occurs. That is, perhaps the timing of plant species B and *Apis* are similar and so the abundance of plant species B is in some ways, a proxy for the abundance of *Apis mellifera*. Finally, the abundance of plant species B may cause an increase in the probability of an interaction between plant species A and *Apis* and vice versa because the occurrence of both attracts *Apis mellifera* more strongly than one or the other and thus makes interactions more likely for both plant species.

**Ranking of *Apis mellifera* preference for various flower species**

When trying to quantify the preference *Apis mellifera* had for the plant species considered in the first model, I ran into the difficulty of finding a single way of comparing a pollinator’s preference. The threshold values and slopes of the logistic regressions for *Gilia capitata*, *Hypericum perforatum*, *Eriophyllum lanatum*, and *Ligusticum grayi* indicate that *Apis* visitation responds to a smaller number of flowers and responds more rapidly to an increase in flowers for *Gilia capitata* than for other species, assuming flowers are counted consistently for all 4 species. However both the abundance at the inflection point and slope of the logistic suffer from the difficulty of comparing the abundance of two flower species. *Gilia capitata* tends to have fewer flowers than *Ligusticum grayi*, for example, as Ligusticum has umbels while the multiple flowers of *Gilia capitata* are so small that in the flower abundance data, they are counted as one flower. The y-intercept was considered as a possible statistics, but this is more likely a measure of the number of MWYs with zero abundance of the target flower and therefore no interactions, as a larger number of such MWYs pulls the curve closer to a y-intercept of zero. I found that probability at the maximum observed abundance for each plant species suffered from the extreme outliers in the abundance data. Finally I considered probability at the mean of the observed abundances, as the median of the observed abundances for all four plant species was zero. This statistic gives an idea of how often an interaction between *Apis mellifera* and the target plant species occurs, as the ranking of the flower species based on these probabilities matches the ranking based on the number of plots *Apis mellifera* interacted with the flower species in. However, it does not give an idea of which *Apis mellifera* might prefer, that is, given both are in a meadow with the same abundance,
which *Apis mellifera* would be more likely visit.

Another option for relative ranking of flower species is to consider the negative coefficients found in the hierarchical model at the meadow network level, or $\mu_\beta$ level. If the abundance of plant B has a negative coefficient in the fit for plant A, but the abundance of plant A does not have a significantly significant coefficient in the fit for plant B, it might suggest a preference for plant B as the presence of plant B decreases the probability *Apis mellifera* visiting plant A but plant A has no such similar effect on plant B. However using this logic and assuming that $A > B$ and $B > C$ implies $A > C$ leads us to internal inconsistencies. An additional complicating factor is the addition of an unattractive option has been shown to change the relative preference *Apis mellifera* has between two other options (Shafir et al., 2002). This would mean that $A > B$ and $B > C$ would not imply $A > C$ since both other preferences were based on the alternatives not captured in the inequality. Thus, some rankings may be guessed as using the $\mu_\beta$ coefficients but those rankings should be conditional on the other species present.

**Conclusion**

The analysis in this paper leaves many openings for future work. A hierarchical model that incorporates the structure of watches within years might better capture the differences between years that the current model with only meadow level coefficients does not capture. Additionally, the basic model could be run with the same 15 species as the hierarchical model to see if similar results are found as in the current hierarchical model.

A similar logistic model could be run for a native pollinator species. *Apis mellifera* is widely used for crop pollination because it will pollinate most flower species but native pollinators may have clearer preferences. *Bombus mixtus*, a native pollinator, was observed visiting 73 different plant species (compared to *Apis mellifera*’s 59) in the interaction data from the H.J. Andrews meadows and so may be a natural choice for a next pollinator.

To investigate the effect of *Apis mellifera* on native pollinators similar model could be run using the abundances of pollinator species in a addition to plant species. Whether or not a pollinator interacts in a plot or meadow could be considered as a function of both the flowers present and other pollinators present. While there may be significant problems with confounding variables, this model may help indicate the effects pollinators have on each other rather than just the effects flowers have on plant species.

**Acknowledgements**

I would like to thank Julia Jones, Tom Dietterich, Alix Gitelman, and Rebecca Hutchinson for their invaluable mentorship for this project. This work would also not have been possible without the Eco-Informatics Summer Institute.

**Literature Cited**


Tables and Figures

Figure 1: 
**Log odds relation between Interaction and Abundance**

The relation between (a) *Gilia capitata*-Apis mellifera interactions and (b) *Eriophyllum lanatum*-Apis mellifera interactions and the abundance of the target flower species in a meadow with the fitted logistic regression overlaid. Abundances were natural log-transformed so the log odds regression curves do not have the usual shape. McFadden’s pseudo $R^2$ is (a) 0.5764 and (b) 0.2301 and the p-value of the abundance coefficients are (a) $p < 0.0001$ and (b)$p < 0.0001$. 

Page 9
Figure 2: Log odds relation between Interaction and Abundance

The relation between the (a) *Hypericum perforatum-Apis mellifera* and (b) *Ligusticum grayi-Apis mellifera* interactions and the abundance of the target flower species in a meadow with the fitted logistic regression overlaid. Abundances were natural log-transformed so the log odds regression curves do not have the usual shape. McFadden’s pseudo $R^2$ is (a) 0.2823 and (b) 0.2238, the coefficients of the abundances are (a) $p < 0.0001$ and (b)$p < 0.0001$, and the p-value of the abundance coefficients are (a) and (b).
Table 1: **Summary from meadow level log odds regression for each plant species** For each plant species, a log odds regression was fitted to the indicator variable for interactions in each meadow as a function of the the abundances of all four plant species for the entire meadow. For each coefficient in a model, the estimate, standard error, z value, and p value are reported. Additionally, a McFadden’s pseudo $R^2$ value is included for a measure of goodness of fit.

### Summary of model fit on meadow level

<table>
<thead>
<tr>
<th>Model: Gilia_captata ~ Gilia_captata_Abundance + Hypericum_perforatum_Abundance + Eriophyllum_lanatum_Abundance + Ligusticum_grayi_Abundance</th>
<th>McFadden’s Pseudo $R^2$: 0.6208826</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept) &amp; -3.1076 &amp; 0.3409 &amp; -9.115 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Gilia_captata_Abundance &amp; 0.0073 &amp; 0.0013 &amp; 5.5418 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Hypericum_perforatum_Abundance &amp; -0.0038 &amp; 0.0032 &amp; -1.1843 &amp; 0.2363</td>
<td></td>
</tr>
<tr>
<td>Eriophyllum_lanatum_Abundance &amp; 0.001 &amp; 3e-04 &amp; 2.9197 &amp; 0.0035</td>
<td></td>
</tr>
<tr>
<td>Ligusticum_grayi_Abundance &amp; 0 &amp; 1e-04 &amp; 0.0374 &amp; 0.9701</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model: Hypericum_perforatum ~ Gilia_captata_Abundance + Hypericum_perforatum_Abundance + Eriophyllum_lanatum_Abundance + Ligusticum_grayi_Abundance</th>
<th>McFadden’s Pseudo $R^2$: 0.3872791</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept) &amp; -3.0459 &amp; 0.3924 &amp; -7.7626 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Gilia_captata_Abundance &amp; -5e-04 &amp; 0.001 &amp; -0.4873 &amp; 0.6261</td>
<td></td>
</tr>
<tr>
<td>Hypericum_perforatum_Abundance &amp; 0.0056 &amp; 0.0012 &amp; 4.7414 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Eriophyllum_lanatum_Abundance &amp; -9e-04 &amp; 0.0013 &amp; -0.651 &amp; 0.5151</td>
<td></td>
</tr>
<tr>
<td>Ligusticum_grayi_Abundance &amp; -1.2631 &amp; 54.1718 &amp; -0.0233 &amp; 0.9814</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model: Eriophyllum_lanatum ~ Gilia_captata_Abundance + Hypericum_perforatum_Abundance + Eriophyllum_lanatum_Abundance + Ligusticum_grayi_Abundance</th>
<th>McFadden’s Pseudo $R^2$: 0.2587804</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept) &amp; -2.2323 &amp; 0.2282 &amp; -9.7808 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Gilia_captata_Abundance &amp; 4e-04 &amp; 2e-04 &amp; 1.8688 &amp; 0.0617</td>
<td></td>
</tr>
<tr>
<td>Hypericum_perforatum_Abundance &amp; -2e-04 &amp; 0.001 &amp; -0.2212 &amp; 0.825</td>
<td></td>
</tr>
<tr>
<td>Eriophyllum_lanatum_Abundance &amp; 0.0025 &amp; 4e-04 &amp; 5.6541 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Ligusticum_grayi_Abundance &amp; 1e-04 &amp; 1e-04 &amp; 0.9981 &amp; 0.3182</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model: Ligusticum_grayi ~ Gilia_captata_Abundance + Hypericum_perforatum_Abundance + Eriophyllum_lanatum_Abundance + Ligusticum_grayi_Abundance</th>
<th>McFadden’s Pseudo $R^2$: 0.255988</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept) &amp; -3.1042 &amp; 0.3292 &amp; -9.4303 &amp; 0</td>
<td></td>
</tr>
<tr>
<td>Gilia_captata_Abundance &amp; 1e-04 &amp; 1e-04 &amp; 0.7427 &amp; 0.4577</td>
<td></td>
</tr>
<tr>
<td>Hypericum_perforatum_Abundance &amp; -10.773 &amp; 625.5042 &amp; -0.0172 &amp; 0.9863</td>
<td></td>
</tr>
<tr>
<td>Eriophyllum_lanatum_Abundance &amp; 2e-04 &amp; 4e-04 &amp; 0.6408 &amp; 0.5216</td>
<td></td>
</tr>
<tr>
<td>Ligusticum_grayi_Abundance &amp; 3e-04 &amp; 1e-04 &amp; 4.4003 &amp; 0</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: **Statistics for ranking Apis mellifera preferences from meadow level analysis** These statistics are calculated using the log odds regression that was fitted to the meadow level data with all four flower abundances as explanatory variables. Note that we would expect a hypothetical preferred plant species to have a low abundance at inflection, high slope at inflection, high probability at maximum observed abundance and high probability at mean observed abundance.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gilia capitata</td>
<td>424.86582</td>
<td>3e-04</td>
<td>0.0428</td>
<td>1</td>
<td>0.44953</td>
</tr>
<tr>
<td>Hypericum perforatum</td>
<td>543.9004</td>
<td>0.00024</td>
<td>0.04539</td>
<td>0.99338</td>
<td>0.06038</td>
</tr>
<tr>
<td>Eriophyllum lanatum</td>
<td>882.24625</td>
<td>0.00022</td>
<td>0.09689</td>
<td>0.99975</td>
<td>0.16509</td>
</tr>
<tr>
<td>Ligusticum grayi</td>
<td>10170.12285</td>
<td>1e-05</td>
<td>0.04294</td>
<td>0.98878</td>
<td>0.05457</td>
</tr>
</tbody>
</table>

Figure 3: **Significant coefficients at forest level** Each target species is a column of the graph and each row is a $\mu_B$ for the abundance of one of the plant species. The tile at the intersection of column A and row B is colored by the mean of the coefficient for abundance B in the model fit for target species A if the mean was statistically different from 0 at a 5% level and grey otherwise. Positive coefficients were expected along this diagonal and negative or not statistically different coefficients elsewhere in the graph.
Figure 4: **Significant coefficients at the meadow level for target species* Gilia capitata.** All of the meadow level $\beta$ coefficients were extracted from the fit with target species *Gilia capitata*. Each column is the $\beta$ for a flower species and each row is a meadow. The tile at the intersection of column A and row B is colored by the mean of the coefficient for abundance A in meadow B in the model fit for target species *Gilia capitata* if the mean was statistically different from 0 at a 5% level and grey otherwise. Positive coefficients were expected in the column for the abundance of the target species, column 1 in this case, and negative or not statistically different coefficients elsewhere in the graph.
Predicting plant-pollinator interactions

Figure 5: **Significant coefficients at the meadow level for target species Agoseris heterophylla**

All of the meadow level $\beta$ coefficients were extracted from the fit with target species *Agoseris heterophylla*. Each column is the $\beta$ for a flower species and each row is a meadow. The tile at the intersection of column A and row B is colored by the mean of the coefficient for abundance A in meadow B in the model fit for target species *Agoseris heterophylla* if the mean was statistically different from 0 at a 5% level and grey otherwise. Positive coefficients were expected in the column for the abundance of the target species, column 7 in this case, and negative or not statistically different coefficients elsewhere in the graph.
Figure 6: Precision-recall curve for the hierarchical model prediction interactions in 2014 data

The precision-recall curve has a range in the precision from 0.4 to 0.6, implying that when predicting the occurrence of *Apis* interactions in the 2014 data using flower abundance data, regardless of the threshold set, approximately 50% of the interactions predicted actually occurred. For thresholds near 1, recall values up to approximately 15% are observed, but for higher levels of recall to be reached, the threshold must be very close to 0. The area under the precision-recall curve is calculated to be 0.5121.