Distribution Patterns of a Large Brazilian Bee

The black orchid bee *Eulaema nigrita* is a species of large bodied Euglossini bees found from Costa Rica through Argentina. This study captured 14443 bees from 178 sites across 72 different locations within Brazil and collected landscape and climate data for each unique location in order to identify patterns of bee distribution related to land use, forest cover, precipitation, temperature, and altitude. Sampling effort and method of capture were also included in the data set.

The purpose of this analysis is to determine and quantify climatic patterns that affect bee distribution. We hypothesize that the strongest determinants of bee abundance will be land use heterogeneity which will have a positive effect on bee abundance, since more diverse landscapes theoretically provide more flowering niches, forest cover which will have a negative effect, since high forest cover would provide fewer flowering niches, and mean precipitation which will have a stabilizing effect preferred above a critical threshold to support flowering plant life and decreasing with higher precipitation.

**Methods**

For this analysis, we first explored the data set and calculated coefficients of variation (CV) between parameters in order to first exclude highly correlated and variables with low explanatory potential. We then fit several generalized linear models to the data set, first creating a global model (where all parameters except for the SU sample identifier were included) and a hypothesized model (including mean average precipitation, land usage heterogeneity, and forest cover as parameters) with Poisson distribution, as the abundance data appeared to not follow a normal distribution but highly clustered at lower value datapoints with only one data point above 1000 bees collected.

> bdat = read.csv("Eulaema.csv", fileEncoding = "Latin1")

> mod = glm(bdat$Eulaema\_nigrita ~ bdat$SA + bdat$method + bdat$effort

+ bdat$altitude + bdat$MAT + bdat$MAP + bdat$forest. + bdat$lu\_het,

family="poisson")

> mod = glm(bdat$Eulaema\_nigrita ~ bdat$MAP + bdat$forest. + bdat$lu\_het,

family="poisson")

However, the outputs of these model summaries showed that the data is overdispersed (the residual divergence is far greater than the degrees of freedom for each of these models; 2681 on 98 degrees of freedom and 17720 on 174 degrees of freedom), therefore the initial models were reconfigured with negative binomial errors to account for this using the MASS package.

> library(MASS)

> mod = glm.nb(bdat$Eulaema\_nigrita ~ bdat$SA + bdat$method + bdat$effort

+ bdat$altitude + bdat$MAT + bdat$MAP + bdat$Tseason + bdat$Pseason

+ bdat$forest. + bdat$lu\_het)

> mod = glm.nb(bdat$Eulaema\_nigrita ~ bdat$MAP + bdat$forest. + bdat$lu\_het)

The outputs of these model summaries showed yielded far lower deviance to degrees of freedom ratios, and the AIC ranks also dramatically decreased, improving the explanatory value of the proposed models (Table 1). However, the addition of the location (SA) significantly improved the rank of the model, so we fit this parameter in a negative binomial model as a fixed effect and in a mixed effect model as a random variable. We also produced a null model with only location as random variable.

> mod = glm.nb(bdat$Eulaema\_nigrita ~ bdat$SA + bdat$MAP + bdat$forest.

+ bdat$lu\_het)

> library(glmmTMB)

> rmod = glmmTMB(data = bdat, Eulaema\_nigrita ~ 1 + (1|SA))

> mem = glmmTMB(data = bdat, Eulaema\_nigrita ~ MAP + forest. + lu\_het

+ (1|SA))

**Results**

Our analysis concluded the bee distribution in this data set is best explained by mean annual precipitation (CV = -0.25), forest cover (CV = - 0.26), land use heterogeneity (CV = 0.09), and the location itself, which combines the effects of all climatic variables specific to each site (CV = -0.23). As hypothesized, precipitation, forest cover, and land use heterogeneity all had high explanatory value, however the addition of the location as a parameter significantly increased the fit of the model, increasing the percent variance explained from 22.2% to 82.4% (Table 1). Location inherently we can expect would have a significant deterministic effect, when considering that all climatic metrics are the same for each discrete location. We also tested a mixed effect model with location as a random effect, however this did not explain the variation (r2 = 0.107).

We excluded altitude and temperature because these variables were highly correlated with one another and with forest cover and seemed to have low explanatory power, as the CVs were low (0.04 and 0.01). Effort was also excluded despite having a high CV, since intuitively and confirmed by the data, more hours spent catching bees will yield more bees caught, and this does little to explain the distribution patterns since effort is not correlated to location.

Land use heterogeneity is a complex variable that may give confusing data if not broken down into components. Pasture makeup we would expect would yield a positive correlation with bee abundance, opposite to forest density.

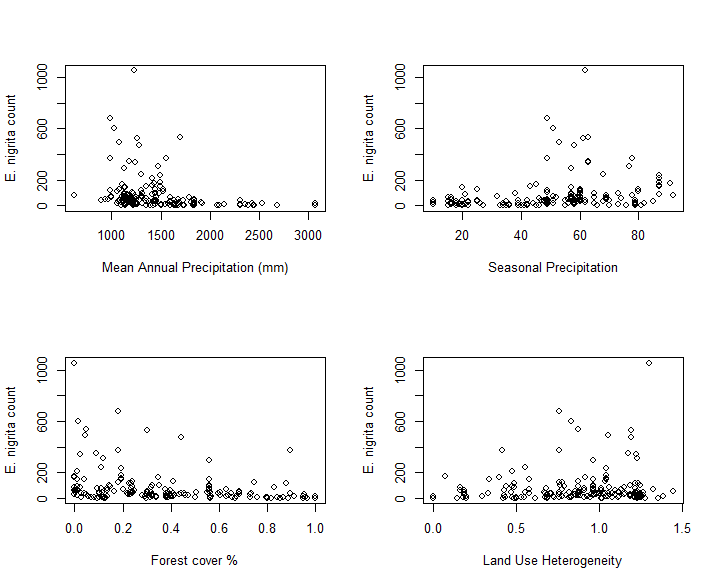
**Figure 1.**

Figure 1: Distribution scatter plots for key parameters. Mean annual precipitation (MAP) showed a negative effect of precipitation on bee abundance however, sampling was largely done in wet tropical biomes and there is only one data point at very low precipitation, which yielded a very low count. Therefore, it is likely that abundance is generally highest around the across site mean (1457 ± 409 mm). Seasonal precipitation was calculated as the coefficient of variation (CV = σ/μ) of monthly precipitation, given as a percent x 100. Although seasonal precipitation seemed to show some correlation with abundance, the effect appeared weakly positive, confounding the negative MAP effect, therefore it was ultimately excluded from our models. Forest cover given as a percentage showed a definite negative effect on bee abundance. Land use heterogeneity was calculated using the Shannon diversity index and showed a positive effect on bee abundance.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **AIC rank** | **Pseudo r2** | **Percent Variance Explained** |
| Global Poisson | 3765.3 | 0.998 | 0.885 |
| Global Negative Binomial | 1687.1 | 0.993 | 0.853 |
| Hypothesized Negative Binomial  MAP, forest cover, land use | 1845.1 | 0.296 | 0.222 |
| Negative Binomial  MAP, forest cover, land use, location | 1704.8 | 0.991 | 0.824 |
| Null Random Effect of location | NA | 1.374e-09 | NA |
| Mixed Effect Model  Deterministic: MAP, forest cover, land use  Random: location | NA | 0.107 | NA |

**Table 1.**

Table 1: Model ranking: negative binomial error model including the hypothesized variables as well as location yielded the best fitting model with explanatory value.

R code can be found in Github repository below under “bee distribution.R”

<https://github.com/mtindall69/bios14>