

Processing and Analysis of Biological Data

Poisson and negative-binomial regression

Maria Niki Chatzantoni

Introduction

The distribution and abundance of bees in Brazil's Atlantic Forest can be affected from both weather conditions and the characteristics of the landscape in which they are active. To study *Eulaema nigrita*, we examine how factors such as mean annual temperature, precipitation and their seasonality, as well as forest cover and land use heterogeneity, contribute to population fluctuations of the species. In WorldClim data, temperature seasonality is expressed as standard deviation (SD), since temperature is measured in a constant unit and absolute variation has biological meaning, while precipitation seasonality is expressed as coefficient of variation (CV), which allows comparison between regions with very different total rainfall levels. In this study, we use GLMs to examine how climate and landscape influence key characteristics of *E. nigrita* abundance. Specifically, we are testing how much weather conditions and their seasonality are related to the abundance of the species, if forest cover and land use heterogeneity increase or decrease their populations and if the method used to capture them affects the observed rates of abundance.

Methods

To investigate how weather conditions and the characteristics of the landscape affect the distribution and abundance of *Eulaema nigrita* in Brazil's Atlantic Forest, we applied GLMs with Poisson distribution. However, Poisson showed overdispersion, therefore we fitted a negative binomial model. For the analysis, we used abundance measurements of bees and a set of climatic and geographic variables (mean annual temperature, rainfall seasonality, forest cover, land use heterogeneity, mean annual precipitation, temperature seasonality and altitude), as well as the collection method. Before fitting the model, we checked for multicollinearity using VIF and excluded variables with high correlation, keeping only those with an independent contribution to the prediction (forest cover, mean annual temperature and precipitation seasonality). For each variable, we calculated the model coefficient and its percentage effect on the predicted abundance, depending on the size of the variable or its category. To account for the different sampling effort, we used $\log(\text{effort})$ as an offset. We plotted the distribution of abundance and effort with histograms and showed the effects of important variables through predictions from the negative binomial model. All analyses and visualizations were carried out in R.

Results

The *Eulaema nigrita* abundance has a right skewed distribution with many zeros and low values, which is typical for count data (Figure 1, A). Sampling effort was used as an offset in the model to express abundance in terms of rate and to adjust for differences in effort between sites (Figure 1, B). The bee abundance showed overdispersion, justifying the use of a Negative Binomial model instead of Poisson.

The selection of final factors was made in two steps to avoid multicollinearity. First, VIF checks showed a high correlation between altitude (VIF ≈ 9.24) and MAT (VIF ≈ 11.61), so altitude was removed, as mean annual temperature has a more direct effect on bee abundance. Next, factors that either provided redundant information (MAP and Tseason in comparison to Pseason and MAT) or had little effect were removed. The sampling method (Net, Traps, or NetTraps) affected abundance by approximately 18% (Table 1), but boxplots (Figure 2) showed that the medians of the methods were very close. Compared to the larger effects of environmental factors, the difference was small, so we removed the method factor from the final model to focus on the main environmental characteristics. Land use heterogeneity had a very small effect ($\approx 1.03\%$) (Table 1). The final model included the most representative and independent environmental factors, which were forest, MAT, and Pseason.

The results of the analysis show that rainfall seasonality (Pseason) is the most important factor (+47.05%) (Table 1), since an increase in this factor leads to a large increase in the predicted abundance (Figure 3, C). Forest cover and mean annual temperature (MAT) have similar effect sizes (-29.75% and +27.05%, respectively) (Table 1) but opposite directions. Forest cover with a negative slope ($\beta < 0$) reduces abundance by 29.75% (Figure 3, A), while an increase in MAT (positive slope $\beta > 0$) increases abundance by 27.05% (Figure 3, B).

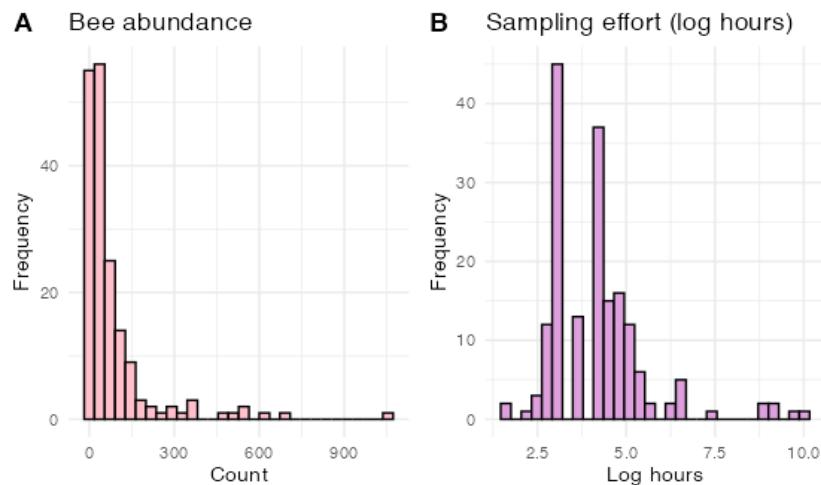


Figure 1: (A) Histogram showing the right skewed distribution of the abundance of *Eulaema nigrita* (B) Histogram illustrating the distribution of sampling effort

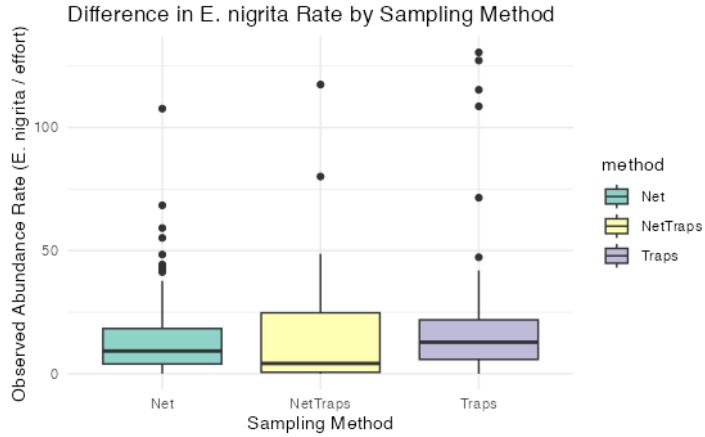


Figure 2: Boxplot that compares the distribution of the abundance rate of *Eulaema nigrita* between the three sampling methods (Net, Traps, NetTraps)

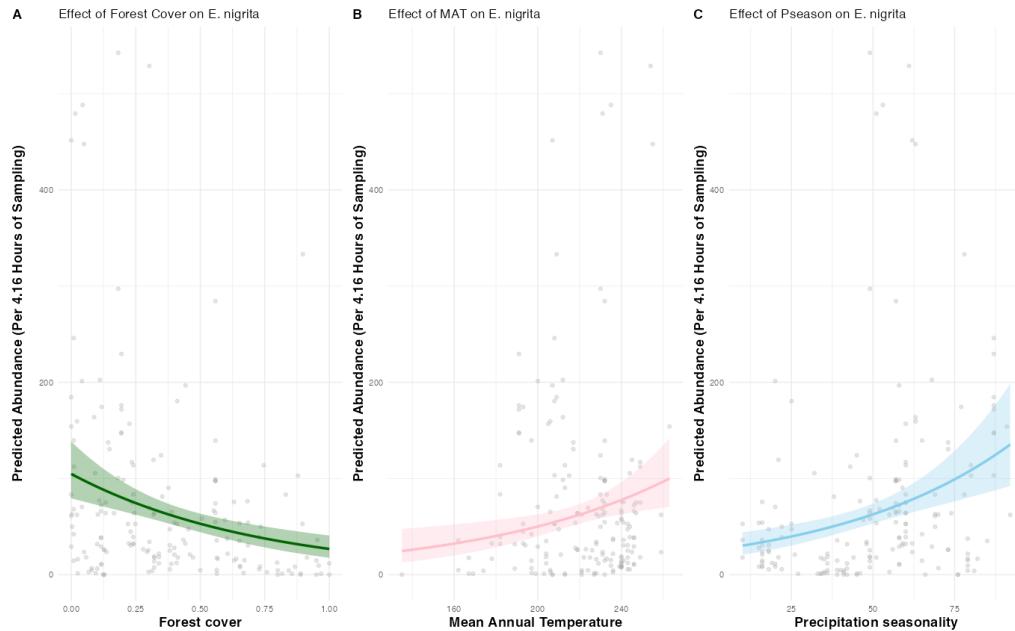


Figure 3: Effect Plots that illustrate the predicted effect of (A) Forest Cover; (B) Mean Annual Temperature (MAT) and (C) Rainfall Seasonality (Pseason) on the abundance rate of *Eulaema nigrita*. The predictions are calculated by keeping all other factors constant (on mean), while the shaded area represents the 95% Confidence Interval

Table 1: Standardized Coefficients and Effect Sizes for Each Predictor

Predictor	$\beta \times SD$ (Standardized Coefficient)	Effect Size (% Change per 1 SD)
Pseason	0.3855	+47.05%
forest	-0.3547	-29.75%
MAT	0.2395	+27.05%
method	(0.1675)	+18.24% (Full Change)
lu_het	0.01021	+1.03% (per SD)

Table 1: Standardized coefficients ($\beta \times SD$) and the corresponding Effect Size for the factors of the model. The Effect Size expresses the percentage change in the predicted abundance rate of *Eulaema nigrita* when each factor increases by one standard deviation (1 SD)

Discussion

The study demonstrates that the abundance of *Eulaema nigrita* in Brazil's Atlantic Forest is strongly influenced by both climatic conditions and landscape characteristics. The results show that rainfall seasonality (Pseason) is the most important factor, with its increase leading to a strong increase in predicted abundance. This positive relationship suggests that the species is favored in areas where there is more variation between dry and wet periods. In terms of ecology, this can be explained by the synchronization of the blooming of many plant species with seasonal rainfall, which creates periods of high resource availability. The bees take advantage of those the periods when there are more resources, showing that they can adapt easily to environments, which are changing during the year.

Similar positive influence is observed for the mean annual temperature (MAT), leading to the outcome that warmer regions favor larger populations of the species. These positive effects on the bee abundance, suggests that *Eulaema nigrita* benefits from climates, where both the temperature and the seasonal precipitation support high nectar and pollen production.

On the other hand, forest cover has a negative influence on the bee abundance, with denser areas of the forest reducing their quantity. This is probably because in more open parts of the forest more sunlight reaches the ground, boosting flowering and making it easier for bees to get to food sources.

In summary, the results show that *Eulaema nigrita* bees seem to like areas with more seasonal rain and higher temperatures, while in compact forests, they are less common, probably because of limited light and fewer flowers.

R Code

The code for this exercise is available at the provided link, inside the report2 folder and named “*bees_chatzantoni.R*”

<https://github.com/marianikichatz/BIOS15/tree/781464458c08a29fc8c4228f8500818ec7fff5c3/report2>