**Introduction:**

In this current investigation, we executed an observational study centered on the morphological dimensions of Tephritis conura across distinct geographic regions situated on the eastern and western peripheries of the Baltic Sea. This analysis primarily examines the influence of regional categorization (west/east) and host plant distinctions (Oleraceum, Heterophyllum) on allopatric and sympatric conditions, particularly emphasizing their interactive impact on the body length of T. conura individuals.

**Methods:**

In the initial phase of our investigation, we undertook an exploratory analysis of the response variable, body length, to gain insights into the underlying data distribution. Therefore, we constructed a histogram to visualize the distribution pattern of body length within the Tephritis population.

Subsequently, we explored the variation in body length distribution across different regions, host plants, and allopatric and sympatric individuals. This step provided additional granularity to our understanding of data distribution, revealing subtle trends within the dataset.

We systematically evaluated various linear and mixed-effect models to discern the most suitable model for our research question. A total of six linear models and one mixed-effect model were assessed, and model selection was guided by the Akaike Information Criterion (AIC).

Having identified the best-fitting model, we analyzed variance (ANOVA) to quantify the statistical differences among Tephritis groups based on their region, host plant, and allopatric or sympatric individual classifications.

Post-model development, we employed data visualization techniques, specifically bar plots and Interaction plots, to effectively communicate the outcomes of our ANOVA analysis. These visualizations act as effective tools to communicate the nuanced insights obtained through our thorough analytical methodology.

**Results:**

Our analysis initiation involved assessing the distribution of the response variable (body length). The histogram revealed a tendency towards a normal distribution, facilitating the application of multiple subsequent analytical tools.

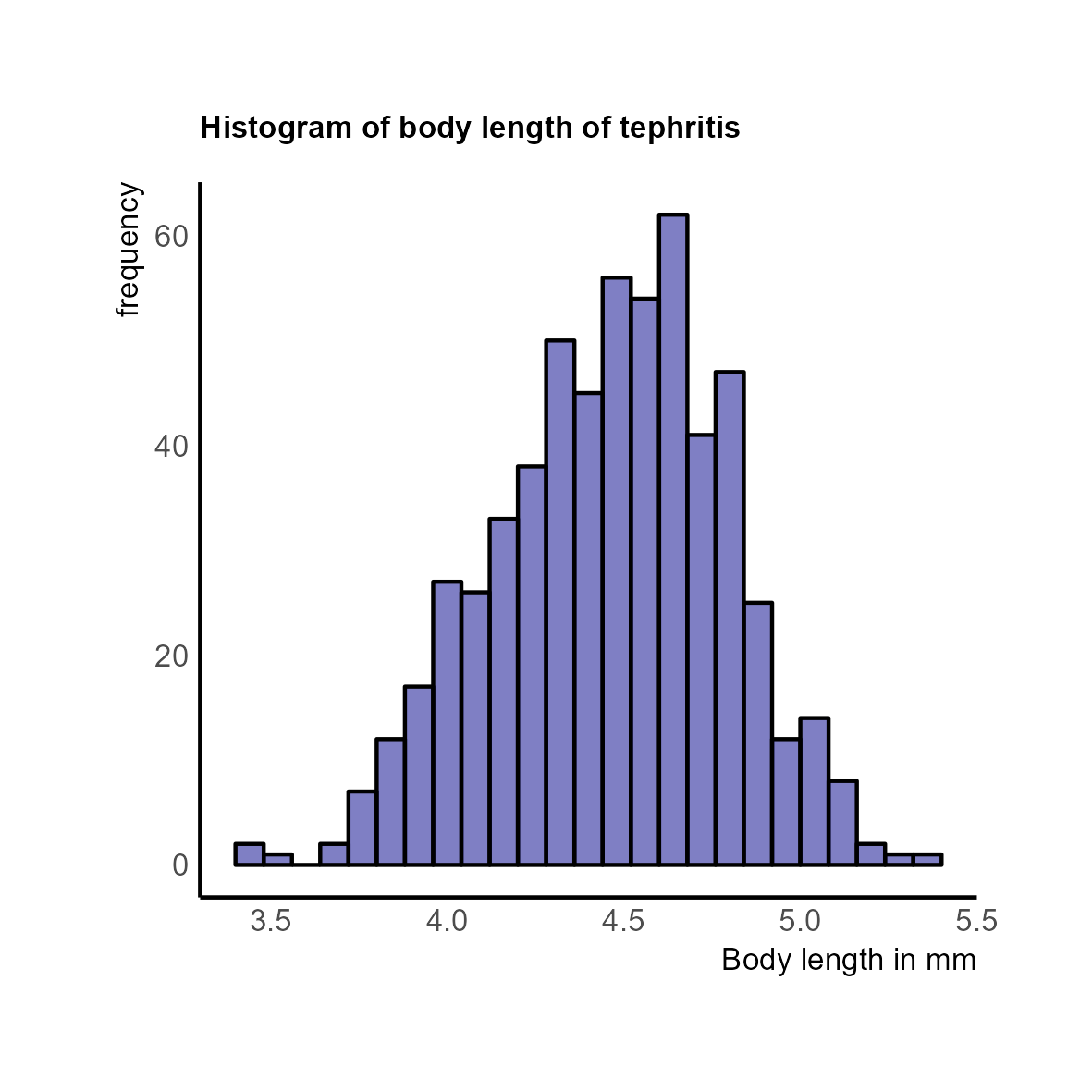


Figure 1: Histogram of body length of tephritis in the dataset.

We initiated the modeling process with a regular GLM using the Poisson family. However, this model exhibited substantial residual deviance of 17042 on 174 degrees of freedom, indicating serious overdispersion. Given this concern, we used a negative binomial distribution similar to the Poisson distribution but with an additional parameter addressing disproportionate variance.

Table 1: General linear model negative binomial distribution (Eulaema nigrita ~ MAP + forest + Pseason)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| General linear model negative binomial distribution (Eulaema nigrita ~ MAP + forest + Pseason) | | | | | | |
| Coefficients: | Estimate | | Std. Error | Z value | P(<|z|) |
| Intercept | 5.348 | | 0.422996 | 12.672 | <2e-16 |
| MAP | -0.00128 | | 0.000214 | -5.966 | 2.43e-9 |
| Forest | -1.09097 | | 0.317824 | -3.433 | 5.98e-4 |
| Pseason | | 0.01995 | 0.004186 | 4.766 | 1.88e-6 |

The intercept represents the estimated log count when all predictor variables are zero (5.35). The negative slope for MAP (-0.0013) and forest cover (-1.09) indicates a decrease in the log count with increasing values of these variables. This translates to a 0.13% negative change in bee abundance for a one-unit change in MAP. The value for forest cover shows us that the abundance of bees in a landscape free of forest is about 109 % higher than in a complete forest. In contrast, the positive slope for precipitation seasonality suggests an increase of 2 % in the abundance of bees for a 1 % increase in precipitation seasonality. The dispersion parameter (0.8368) aligns with the appropriate fit of the negative binomial distribution.

After looking at the combined model incorporating all three predictor variables demonstrating their collective impact on the data, we created two negative binomial GLMs, one incorporating forest cover and mean annual precipitation and the other incorporating precipitation seasonality and mean annual precipitation. These models aimed to provide a nuanced understanding of the individual and combined effects of the predictor variables on Eulaema nigrita abundances.

Table 2: General linear model negative binomial distribution (Eulaema nigrita ~ MAP + forest)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| General linear model negative binomial distribution (Eulaema nigrita ~ MAP + forest) | | | | | | |
| Coefficients: | Estimate | | Std. Error | Z value | P(<|z|) |
| Intercept | 6.67929 | | 0.32959 | 20.265 | <2e-16 |
| MAP | -0.0014 | | 0.00022 | -6.242 | 4.32e-10 |
| Forest | | -1.3318 | 0.31706 | -4.137 | 3.51e-05 |

Table 3: General linear model negative binomial distribution (Eulaema nigrita ~ MAP + Pseason)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| General linear model negative binomial distribution (Eulaema nigrita ~ MAP + Pseason) | | | | | | |
| Coefficients: | Estimate | | Std. Error | Z value | P(<|z|) |
| Intercept | 5.03358 | | 0.412482 | 12.203 | <2e-16 |
| MAP | -0.00145 | | 0.000216 | -6.696 | 2.14e-11 |
| Pseason | | 0.02384 | 0.004079 | 5.844 | 5.1e-09 |

The comparative analysis reveals a similarity in the slope of the mean annual precipitation (MAP) factor across both models. Additionally, it is noteworthy that forest cover continues to negatively influence the logarithmic count, while precipitation seasonality manifests a positive impact. Notably, in the individual effect models, these influences exhibit an increment of approximately 20% when contrasted with the comprehensive model encompassing all three predictor variables. Both models exhibit a dispersion parameter of over 0.75, affirming the congruence of the negative binomial distribution model with the observed data.

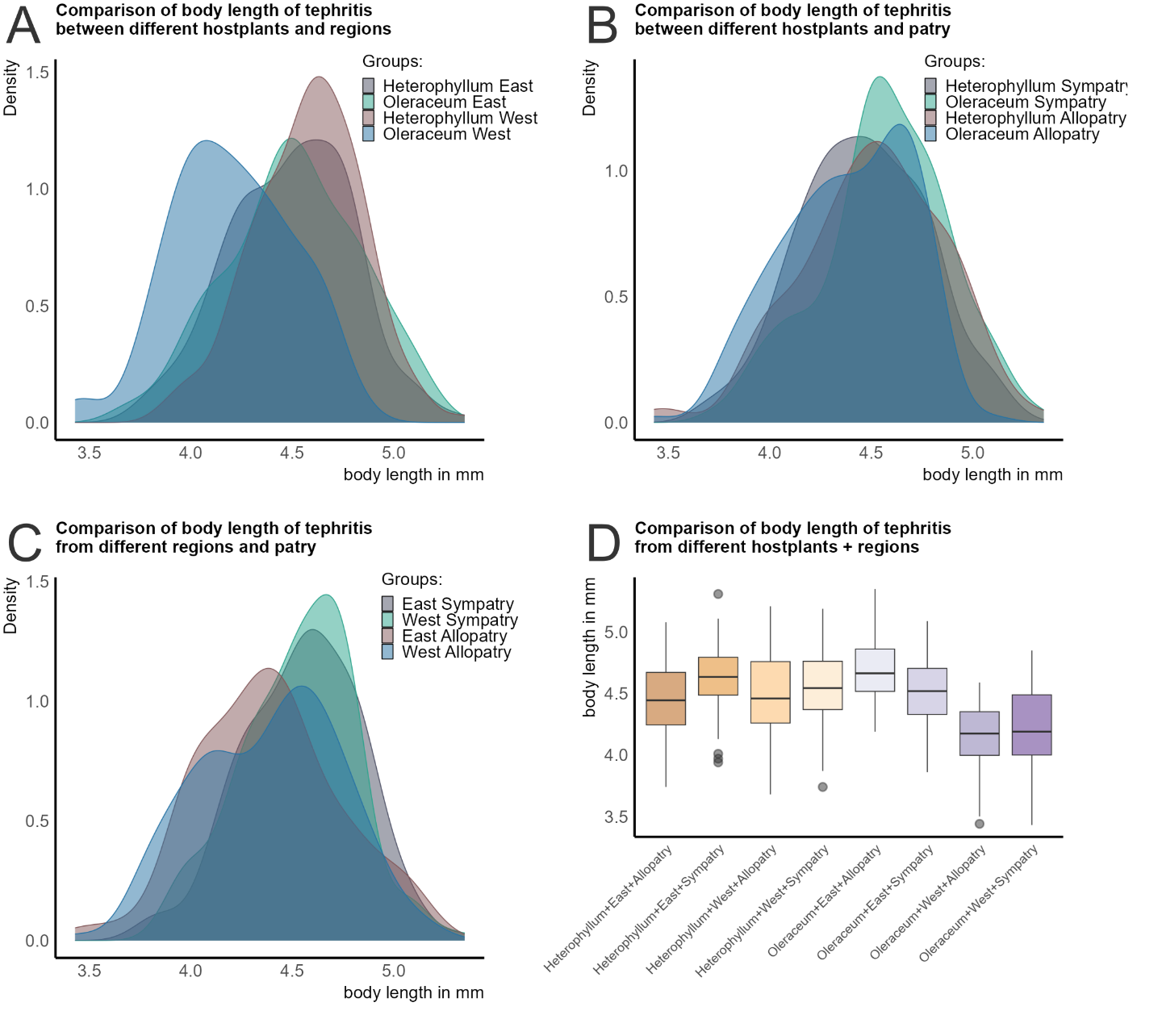


Figure 2: Impact of Forest cover and mean annual precipitation on El. nigrita abundance

The graphical representation depicts a positive correlation between the abundance of bees and a diminishing forest cover. Furthermore, it is observed that a reduced rate of rainfall (MAP=Mean–SD) exerts a favorable influence on the observed abundance of bees, while an increased rate of rainfall(MAP= Mean+SD) decreases the abundance of bees.

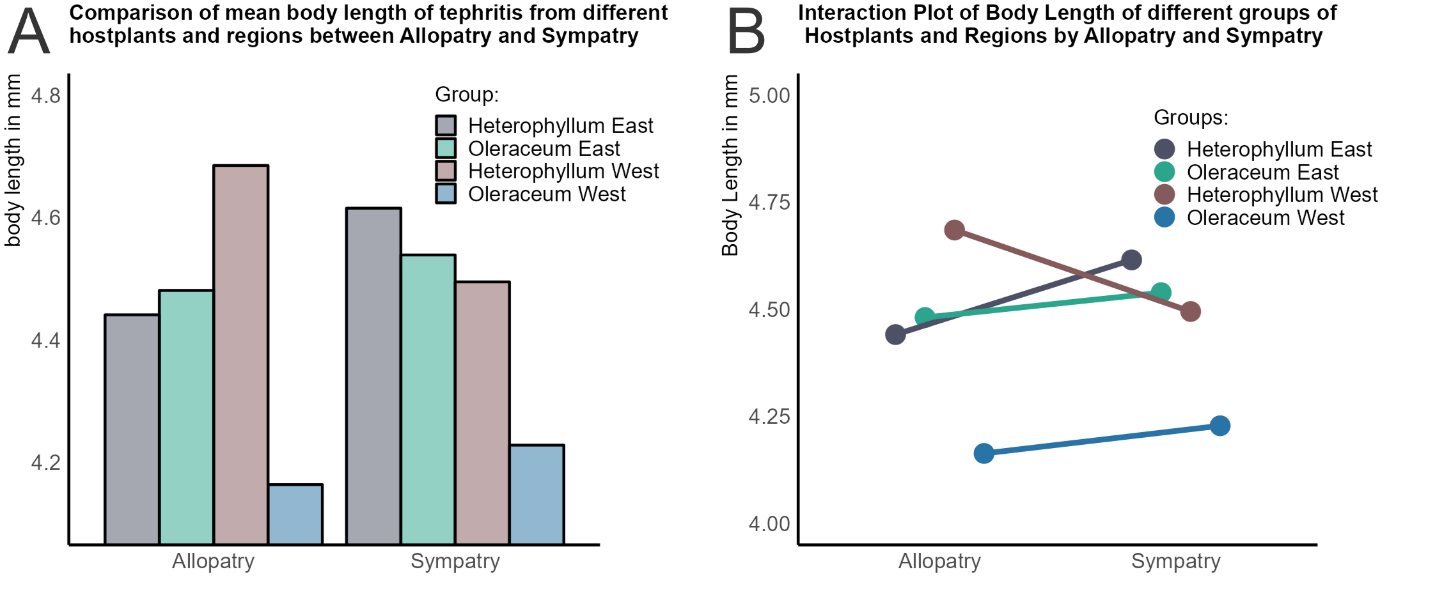


Figure 3: Impact of precipitation seasonality and mean annual precipitation on El. nigrita abundance

The graphical presentation reveals a similar Mean Annual Precipitation (MAP) impact. However, notable distinctions arise as the effect of precipitation seasonality surpasses that of forest cover. Additionally, it is noteworthy that the positive effect indicates that heightened precipitation seasonality corresponds to an increased abundance of bees.

**Conclusion:**

Our investigation shows how precipitation factors and forest cover influence bee abundance. The analysis unveiled a negative impact of forest cover and mean annual precipitation (MAP) on the abundance of bees. Specifically, an elevated proportion of forest cover and increased mean annual precipitation exhibited an inverse correlation with bee sightings. A plausible conclusion suggests that heightened precipitation levels may impede bee flight, as rain conditions challenge their aerial activities. Furthermore, the preference for lower forest cover implies that typical forest tree environments are less conducive for bees.

In contrast, a positive impact was discerned in precipitation seasonality, where a higher seasonality around the average rainfall rate positively influenced bee sightings. We suggest that during months with significantly reduced rainfall, bees tend to engage in more frequent flight activities. The intricate relationships identified in this study contribute to a nuanced understanding of how precipitation factors and land cover jointly shape the abundance patterns of bee populations.