Anna’s Hummingbirds Spatial Distribution Analysis

Andrew Sarracini

Han Chen

Nayeli Montiel

The University of British Columbia

DATA 589 Spatial Statistics

Professor: Michel Noonan

April 28th, 2024

Introduction

Anna’s Hummingbird (*Calypte anna*) is a species of hummingbird most commonly observed in the Pacific Northwestern and coastal regions of North America. Distinctive for their small (10-11 cm) size, iridescent emerald bodies and incredibly fast movement (up to 80 kph!) [1] during a dive, Anna’s Hummingbirds are as well-known as they are appreciated by gardeners and bird watchers alike. Classified as “Least Concern” on the world-wide accepted species endangerment scale, the “Red List”; there are an estimated 9.6 million (and increasing) Anna’s Hummingbirds currently in existence [2]. The data gathered for this report was derived from the Global Biodiversity Information Facility (GBIF), a global organization dedicated to open-data policies regarding access to biodiversity-related information [3]. The data pulled had a range of 74 years (1950 - 2024), with the majority of occurrences in the United States and Canada.

From the beginning of our analysis, we identified several interesting dynamics surrounding the behavior, localization, and migration patterns of Anna’s Hummingbird, we are most interested in these key areas:

- Armed with the knowledge that most hummingbird variants tend to be attracted to human-constructed sugar water feeders, what is the relationship between Anna’s Hummingbirds and the Human Footprint Index?
- Is there a relationship between Anna’s Hummingbirds and spatial elevation? If so, is this more related to temperature or the existence of a particular food source?
- Given the incredible movement that hummingbirds can maintain, do Anna’s Hummingbirds have any natural predators? If so, do they demonstrate avoidant patterns that can be mapped to the Forest Density covariate?

Methodology

The original species dataset is composed of 291,431 and 223 variables. As a preprocessing step we filtered the data by country and province to consider the locations of hummingbirds in British Columbia (BC), Canada. Then, we consider only observations in 2024. Resulting in 302 observations and two columns corresponding to raw latitude and longitude values. It's important to mention that Anna's Hummingbird breeds from December through May and occupies its non-breeding range from July through December [4], which explains the lower number of observations in the year of study. Moreover, all these observations were published by *iNaturalist.org*, an online social network of people sharing biodiversity information to help each other learn about nature.

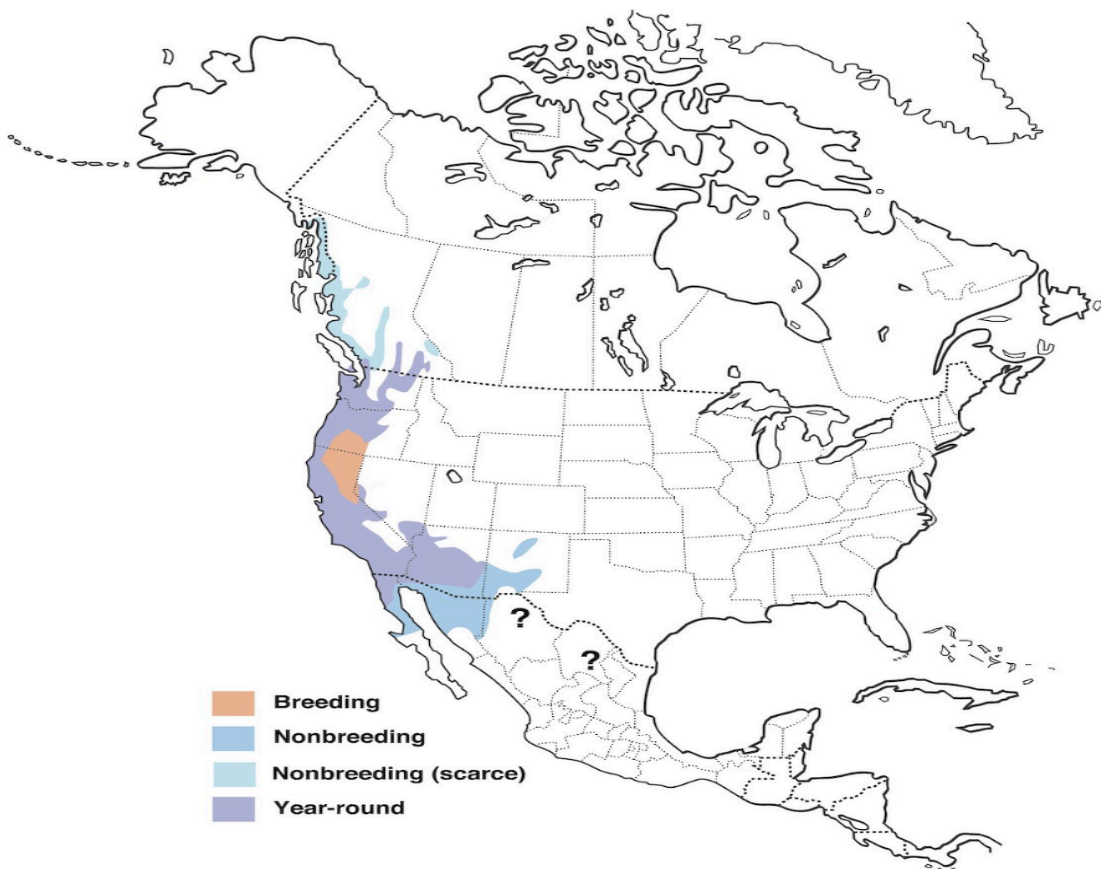

Figure 1. Range of Anna's Hummingbirds

For the data point locations, we consider a transformation from the decimal Latitude and Longitude, to BC Albers projection [5]. Once in the correct range, we created a point pattern object, a set of points in a two dimensional space in a bounded window, which represents the province of British Columbia. During this process, 26 points were rejected because they were lying outside the specified window. Additionally, 16 duplicate points were found, perhaps due to the occurrence of multiple records in the same place and time, provided by different users. The final data is therefore composed of 260 points.

Covariate Definitions

Four environmental covariates were included to analyze their relationship with Anna’s Hummingbird locations: elevation in meters (Elevation), forest cover in percentage (Forest), human footprint index (HFI) and distance to water in meters (Dist_water).

Elevation refers to the height of a geographic location above a fixed reference point, in this case, sea level. Specifically, this covariate augments our understanding about how the altitude of a location influences the presence of Anna’s Hummingbirds in BC. According to Wikipedia [6], 75 percent of the province is mountainous (more than 1,000 meters above sea level), this fact is supported by the data as the median elevation in BC is 1,096.86 meters above the sea level. Meanwhile, the median elevation where Anna’s Hummingbirds are located is around 45.48 meters above sea level, as shown in Figure 2.

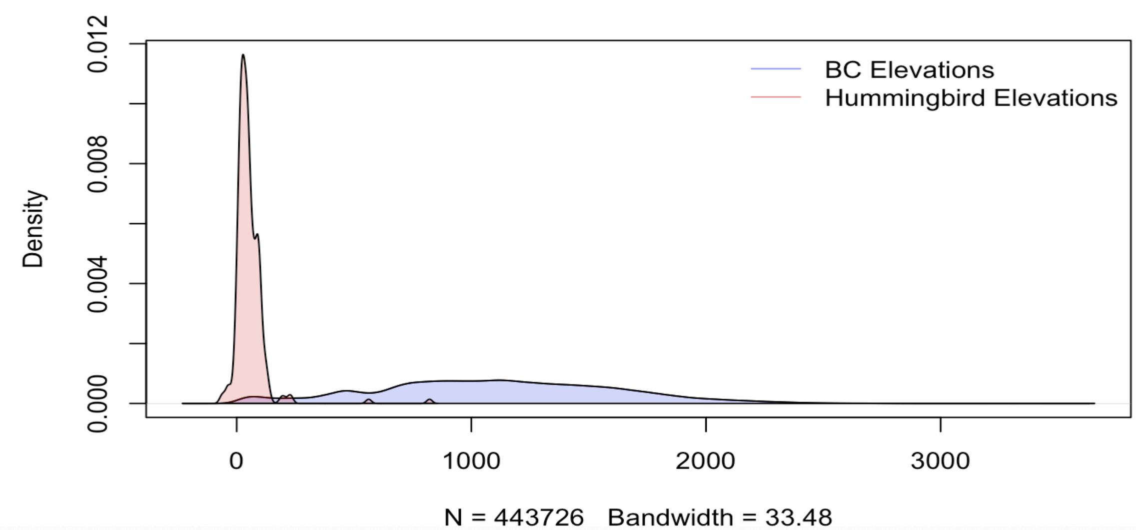

Figure 2. Kernel Density for elevations in BC versus elevations of Anna’s hummingbirds

Forest cover is the area covered by forest, including both natural and planted forest. According to our data, 50.16 percent of the land in BC is considered forested and the median forest cover for points where hummingbirds were observed is 9.01 percent, see Figure 3.

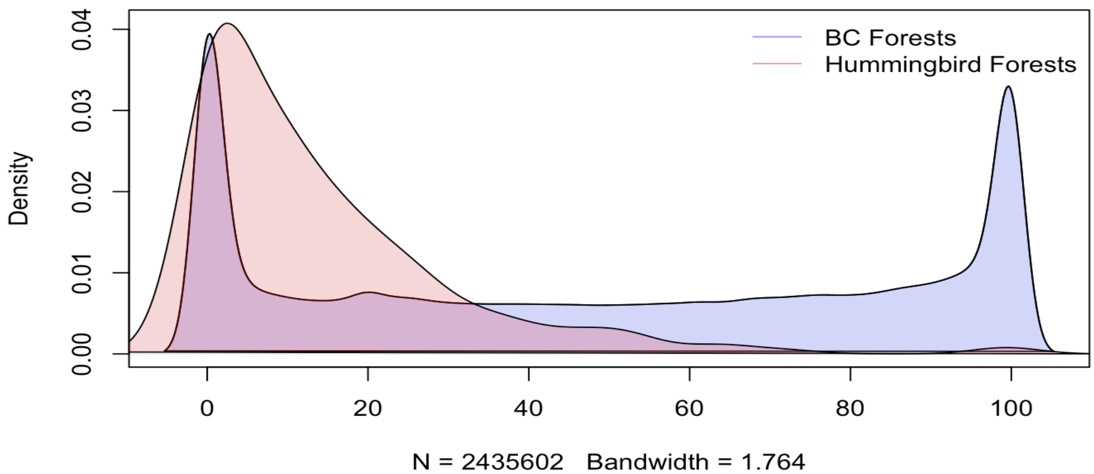

Figure 3. Kernel Density for forest cover in BC versus forest cover of Anna's hummingbirds

Next, HFI, is a metric that defines the impact of Human activity on the Earth's surface, measured on a scale from 0 to 1 (0 being minimal influence). This covariate can be used to make conclusions about the behavior of Anna's Hummingbirds, which could have a ripple effect on their habitat and migration patterns, as they are commonly observed in rural and suburban settings [7]. This is supported by the data, as the median value for areas where hummingbirds were observed is 0.82, see Figure 4.

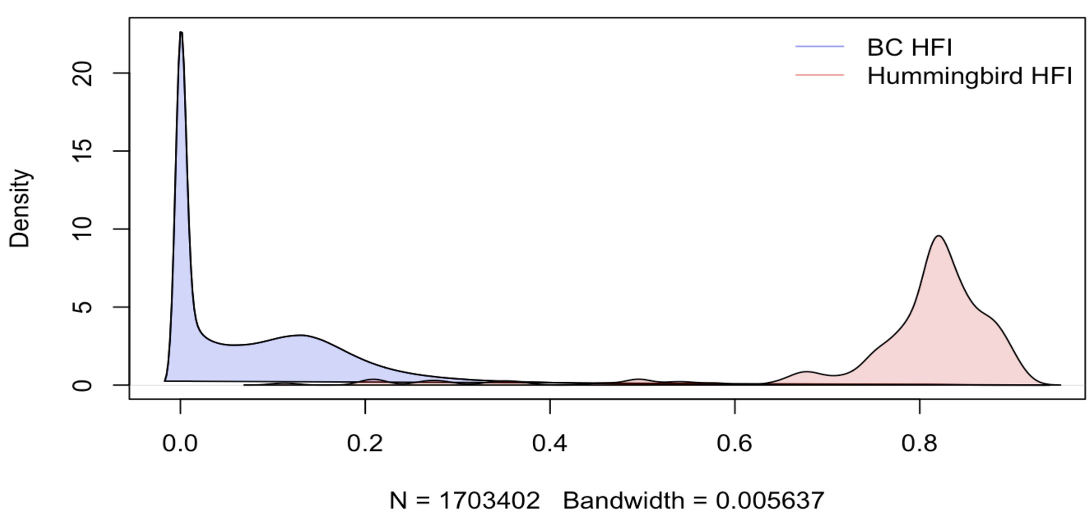

Figure 4. Kernel Density for HFI cover in BC versus HFI of Anna's hummingbirds

Finally, distance to water measures a given point’s proximity to water bodies such as rivers, lakes or oceans. This covariate has the potential to be informative, as relative distance to water sources may be an important factor in habitat selection and better understanding the feeding range of this species. In BC, the median distance to water is approximately 1,095.06 meters. However, the median distance to water at where Anna's Hummingbirds were observed is 536.58 meters, see Figure 5.

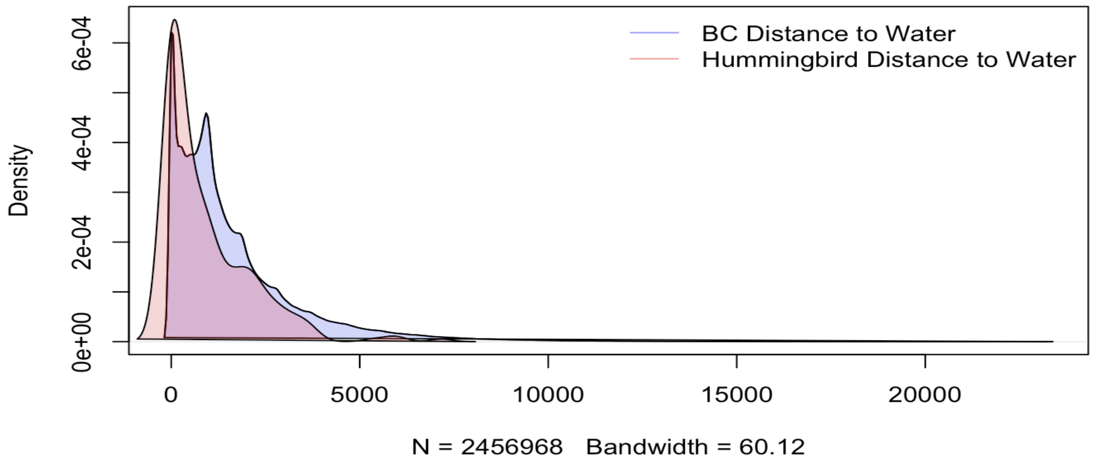

Figure 5. Kernel Density for distance to water cover in BC versus distance to water of Anna’s hummingbirds.

To analyze the locations of Anna's Hummingbirds, we began by visualizing the data and computing summary statistics. Next, we model the point pattern using the Poisson point process (ppp) model and estimate the relationships between points and covariates, through first and second moments descriptive statistics. We also employed model selection and goodness-of-fit tests to evaluate our models, refining them until we found the most appropriate model to explain the intensity of Anna's Hummingbirds in British Columbia.

Figure 6. Anna’s hummingbirds point pattern data in 2024

Experiment

First Moment Descriptive Statistics

We initially explored the intensity of Anna’s Hummingbirds across BC and discovered an inhomogeneous relationship. This result is evidenced by the Chi-squared test of complete spatial randomness (CSR), using quadrat counts showed a p-value smaller than 2.2e-16. This is a strong outcome that refutes the null assumption of a homogeneous process.

Next, we used kernel estimation to delve into the lambda(u) function with "Likelihood Cross Validation" to set the bandwidth. As Figure 7 shows, our estimated result indicates that most points are in the bottom left corner of the observational window, with few in the other areas, suggesting a significantly unbalanced distribution, and perhaps even clustering.

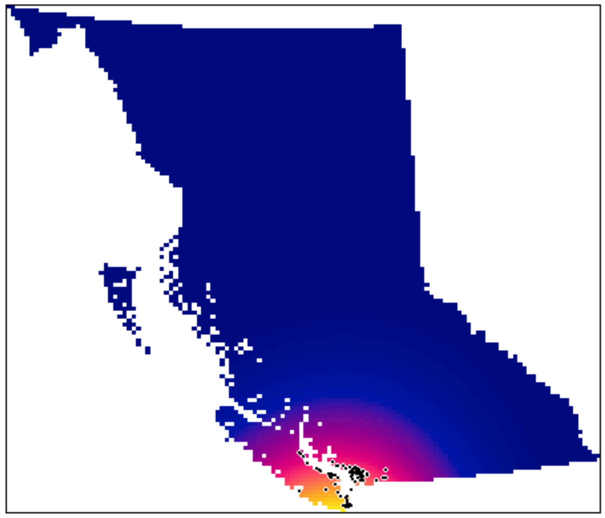

Figure 7. Kernel Estimate with Likelihood Cross Validation Bandwidth

To enhance the objectivity of our results, we also applied hotspot analysis to identify areas of concentrated activity, offering valuable insights into spatial processes. As illustrated in Figure 8, there are two significant intensities highlighted as red squares, while virtually all other aspects of the plot show an intensity of 0.

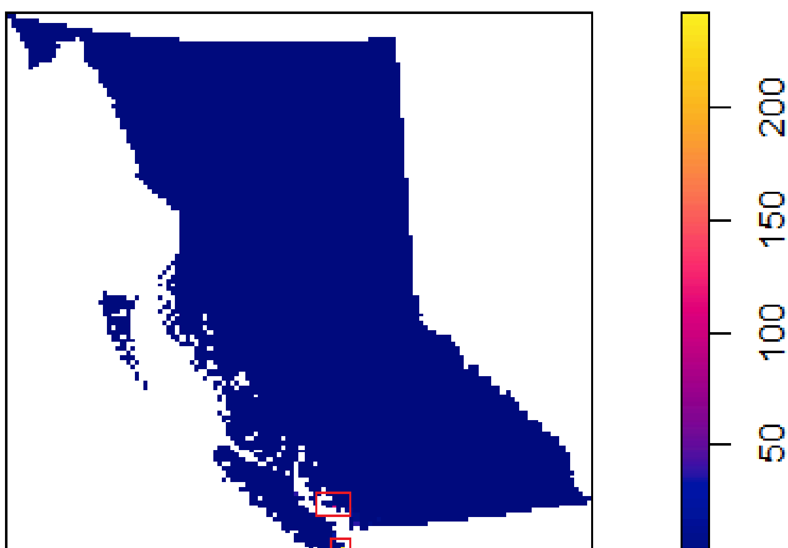

Figure 8. Hot Spot Analysis

Second Moment Descriptive Statistics

After computing first moment descriptive statistics, we revealed the inhomogeneous nature of the data, but the cause of this relationship was still unknown. To determine whether this was due to the point process being heterogeneous or the points not being independent, we conducted further analysis to understand the relationship between the individual points.

As shown in Figure 9, we established a 99% bootstrapped confidence interval to demonstrate this point. Below 7500m, there is no overlap between the observed K function and the estimated K function, indicating a significant clustering of points. However, above 7500m, the points may exhibit some independence, though the relationship is still not very strong.

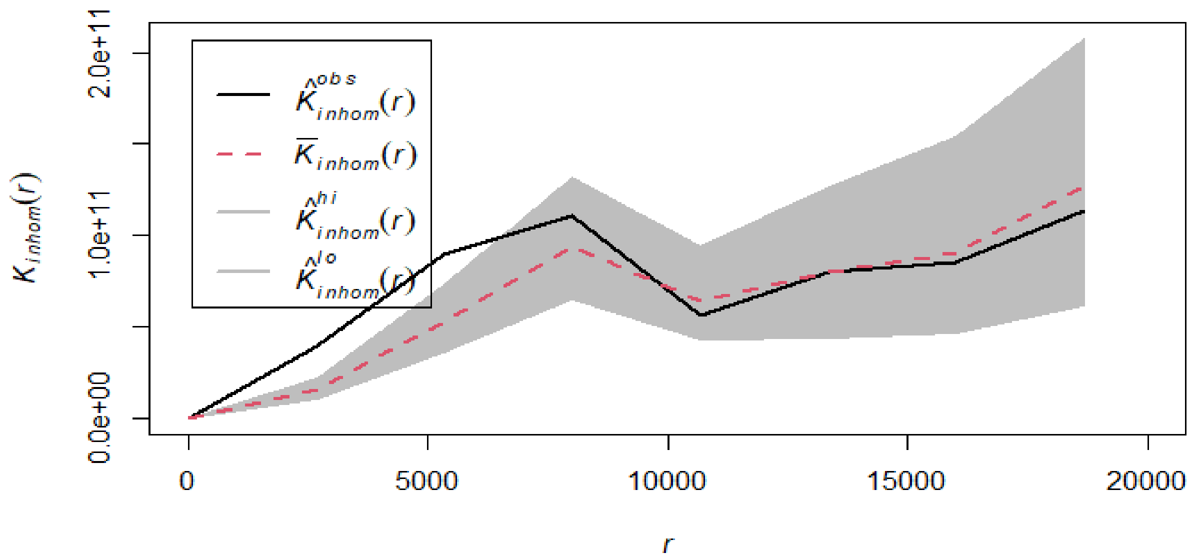

Figure 9. Zoom-in Inhomogeneous Ripley's K

The Inhomogeneous Ripley's K calculates cumulative points, providing limited information on the behavior of the process. Therefore, we further analyze the pair-correlation function. As depicted in Figure 10, despite setting α to 0.01, there is no overlap observed. This suggests that, according to the pair correlation function (pcf), the points are independent. This finding contradicts the results obtained from the K function.

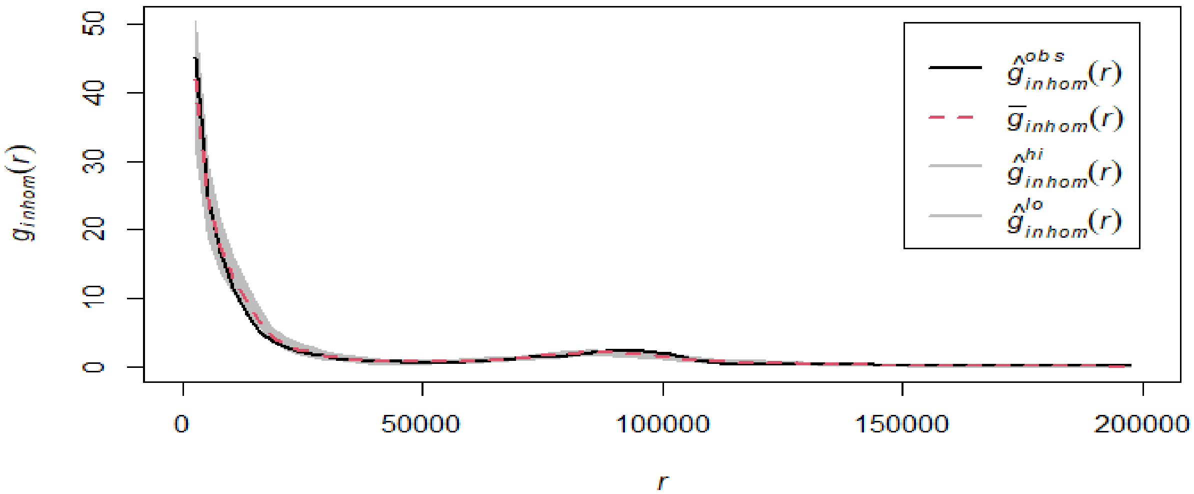

Figure 10. Zoom-in Inhomogeneous Pair Correlation Function

<u>Modeling</u>

Before fitting models to the Hummingbird data, we aim to determine whether the intensity depends on provided covariates: elevation, forest, distance to water, and HFI in BC, as well as the underlying nature of these relationships. As depicted in Figures 11-14, non-linear relationships are evident in both Elevation and Forest covariates. Additionally, the relationships between Distance to Water and Human Footprint Index and our data points appear to be more complex.

- Elevation exhibits a negative relationship with intensity, particularly noticeable up to elevations of 25m, indicating a decrease in the number of hummingbirds as elevation increases.
- Forest cover displays a right-skewed distribution, with intensity increasing up to 10% forest cover and decreasing thereafter.
- Distance to water demonstrates a downward trend, with the highest concentration of birds observed within 0 to 5000m of water.
- HFI exhibits a positive relationship with intensity, most clearly observed at extreme HFI values. Furthermore, it is surprising the curve doesn't drop more at extreme HFI values with less data samples.

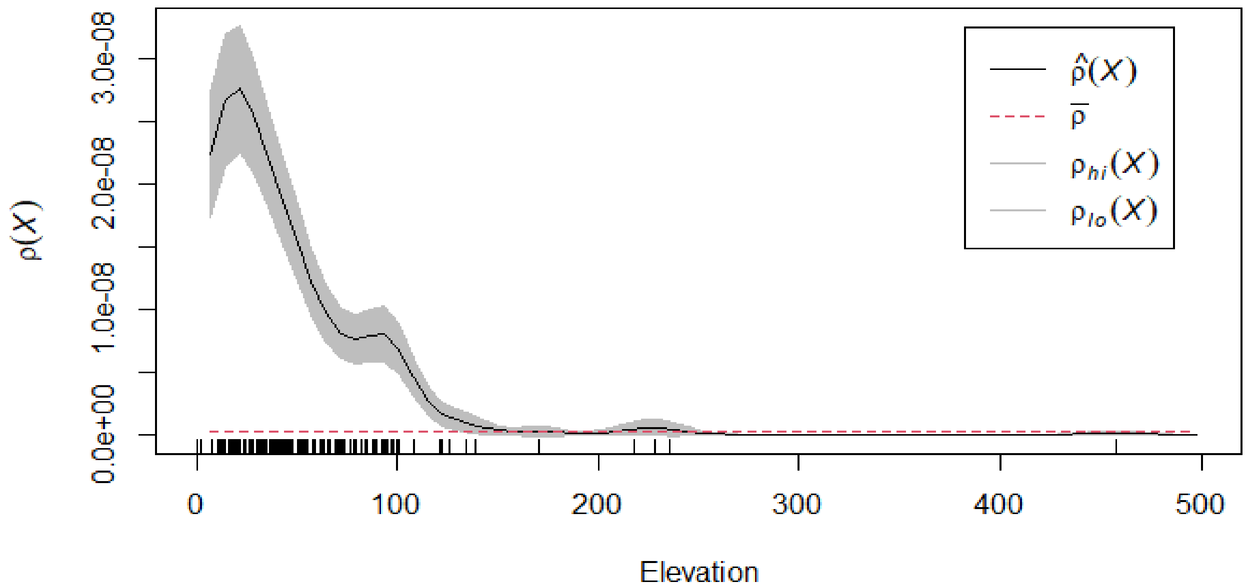

Figure 11. Relationship between Anna’s Hummingbird and Elevation

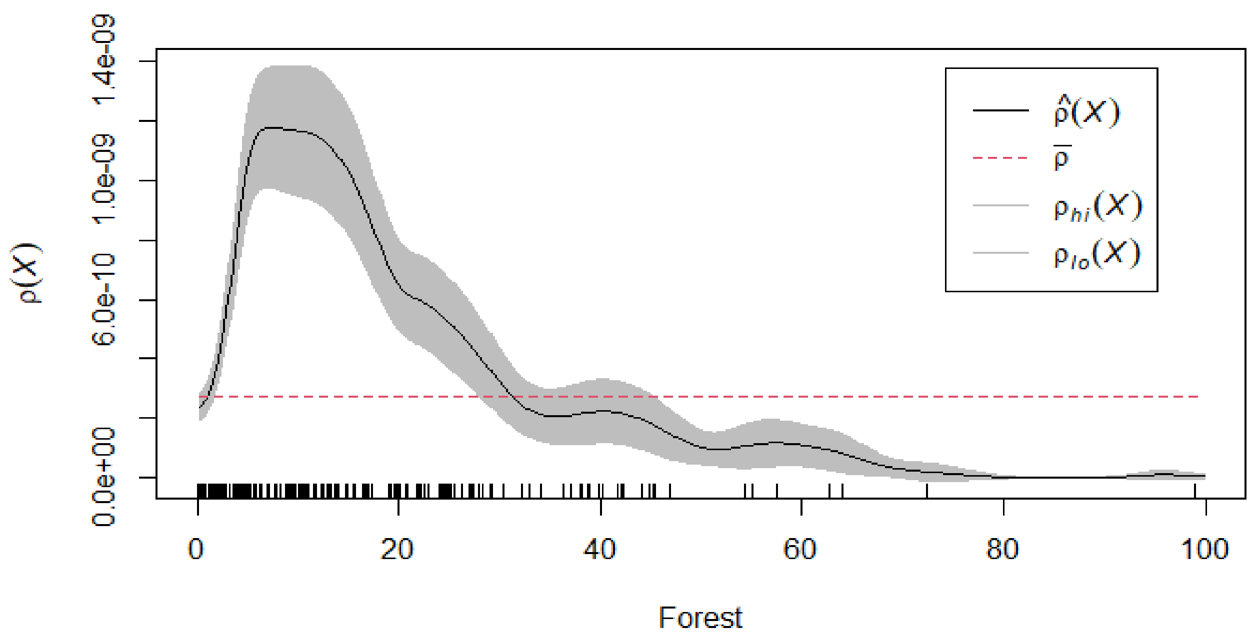

Figure 12. Relationship between Anna’s Hummingbirds and Forest Cover

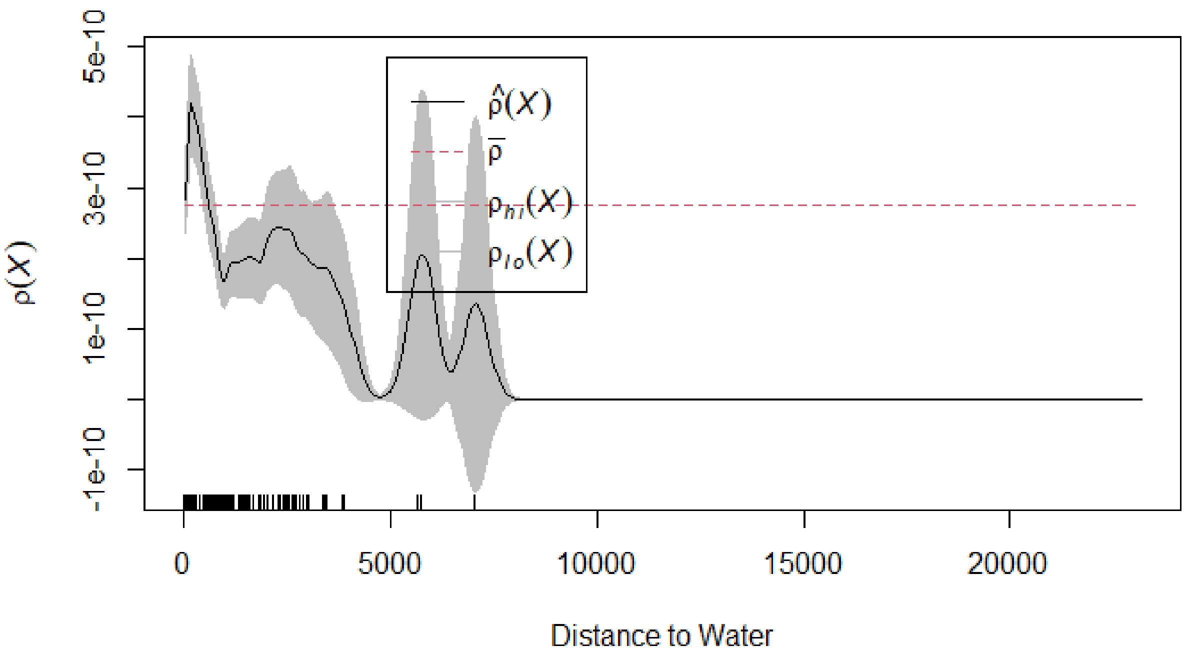

Figure 13. Relationship between Anna's Hummingbirds and Distance to water

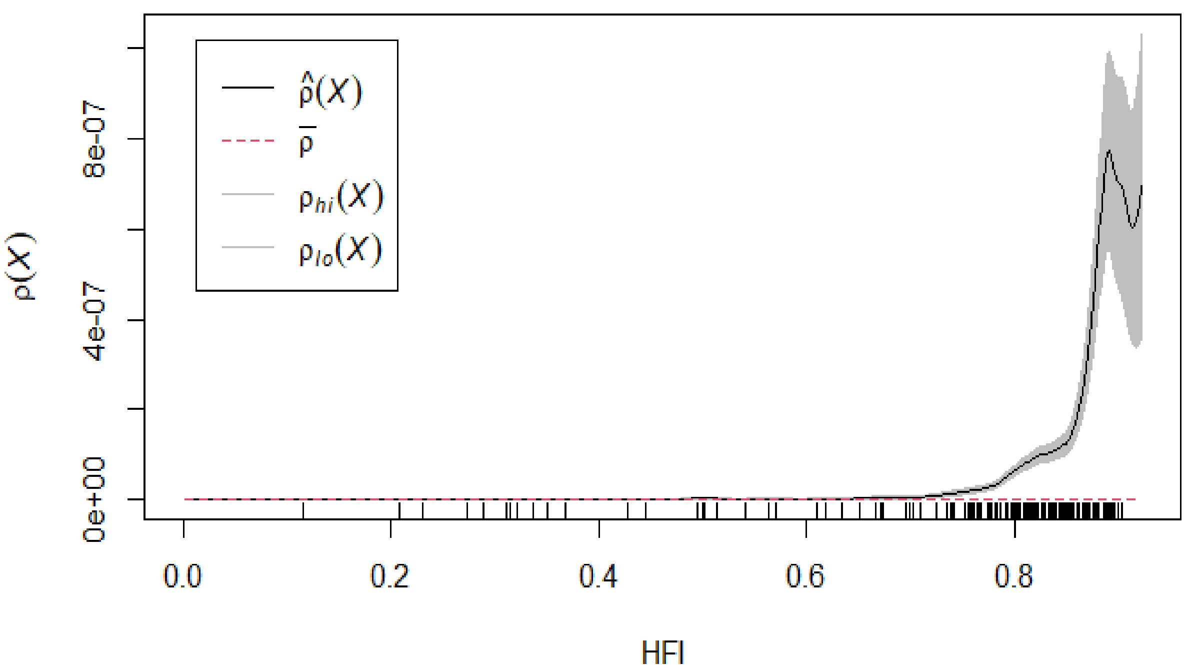

Figure 14. Relationship between Anna's Hummingbird and Human Footprint Index (HFI)

We started with a null model as our baseline and experimented with a linear model, several quadratic models, and Generalized Additive Models (GAM). In the following section, we further explain the model selection and model evaluation we employed to determine the following GAM model as our best fit. The covariates elevation, forest and HFI are modeled using natural spline functions with three degrees of freedom, meanwhile the distance of water is split into eight degrees of freedom. These coefficients indicate the change in log intensity for a one-unit increase in elevation.

- Intercept:
  - The estimated log intensity is -15.44. This means that when all other predictors are zero, the estimated intensity is 3.77×(10^−7).
- Elevation:
  - bs(Elevation, df = 3)1: Estimated coefficient is -21.7766176.
  - bs(Elevation, df = 3)2: Estimated coefficient is -30.2661224.
  - bs(Elevation, df = 3)3: Estimated coefficient is -7.8824194.
- Forest:
  - bs(Forest, df = 3)1: Estimated coefficient is 0.6514961.
  - bs(Forest, df = 3)2: Estimated coefficient is -7.2018404.
  - bs(Forest, df = 3)3: Estimated coefficient is -4.0651657.
- HFI_change:
  - Two of the three degrees of freedom have estimated coefficients of 0, with the final degree of freedom being near-zero. However, this is the only covariate with a non-negative coefficient.
- Dist_Water:
  - Dist_Water is split into eight degrees of freedom.
  - Coefficients range widely, some even exceeding 100, indicating a potentially significant impact of Dist_Water on intensity.

Comparison and Results

<u>Model Selection</u>

The model selection process aimed to identify the best-fitting model for predicting Anna’s hummingbird locations based on four environmental covariates. Starting with a null model, subsequent models including linear and quadratic terms for elevation, forest cover, HFI, and distance to water were tested. While quadratic models improved upon the linear model, the introduction of generalized additive models (GAMs) with spline terms provided an even better fit. Among the GAMs tested, Model 9, incorporating cubic splines for distance to water and quadratic terms for elevation, forest cover, and HFI, exhibited the lowest Akaike Information Criterion (AIC), indicating its superiority. Therefore, Model 10, the GAM with splines for all four covariates, was chosen as the best model for predicting Anna's Hummingbird locations. The following Table 1 summarizes the results for the ten models.

Model Number	Model Name	AIC
1	Null (only intercept)	120,878.44
2	Linear (with 4 covariates)	11,242.56
3	Quadratic elevation (one covariate)	11,199.09
4	Quadratic forest cover (one covariate)	11,901.89
5	Quadratic distance to water (one covariate)	12,050.73
6	Quadratic HFI (one covariate)	12,098.81
7	Quadratic elevation, forest cover and HFI (with 3 covariates)	9,631.574
8	Quadratic full (with 4 covariates)	9,630.002
9	GAM cubic spline for distance to water with 8 df and quadratic terms for elevation, forest cover and HFI (with 4 covariates)	9,620.369
10	GAM cubic spline for elevation, forest cover and HFI all with 3 degrees of freedom and distance to water with 8 df (with 4 covariates)	9,612.959

Table 1. AIC for Model Selection

Likelihood Ratio Test

We conducted likelihood ratio tests to compare pairs of nested models.

- The comparison between the null model (intercept only), Model 1, and the linear model with four covariates, Model 2, yielded a p-value of 2.2e-16, indicating significant evidence against the null model in favor of the linear model.
- Comparisons between the linear model, Model 2, and each quadratic model (elevation, forest cover, distance to water, and HFI), Models 3 to 6, resulted in tiny p-values (all < 0.05), suggesting that the quadratic models with one covariate are better fits than the linear model.
- Similarly, the comparison between the linear model (Model 2) and the quadratic model with four covariates (Model 8) also yielded a tiny p-value (2.2e-16), indicating the superiority of the quadratic model with four covariates.
- When comparing the quadratic model with three covariates (Model 7) to the quadratic model with four covariates (Model 8), the p-value was 0.06169, above the significance level of 0.05. This suggests that the quadratic model with three covariates is a better fit for our point pattern, supported by the non-linear pattern observed in the distance to water covariate.
- The likelihood ratio test comparing the linear model (Model 2) to the GAM model with 4 covariates and spline in distance to water (Model 9) resulted in a tiny p-value of 2.2e-16, indicating the superiority of the GAM model.
- Comparing the quadratic model with three covariates (Model 7) to the GAM model with 4 covariates and spline in distance to water (Model 9) yielded a small p-value of 0.001411 (<0.05=alpha), suggesting the GAM model is a better fit.
- Similarly, comparing the GAM model with 4 covariates and spline in distance to water (Model 9) to the GAM model with splines in the four covariates (Model 10) resulted in a small p-value of 0.00383, indicating the GAM model with four covariates is the best fit.

Overall, both the AIC criteria and likelihood ratio tests support the more complex GAM model with splines in the four covariates (Model 10) as the best one given out observations on Anna's Hummingbird locations.

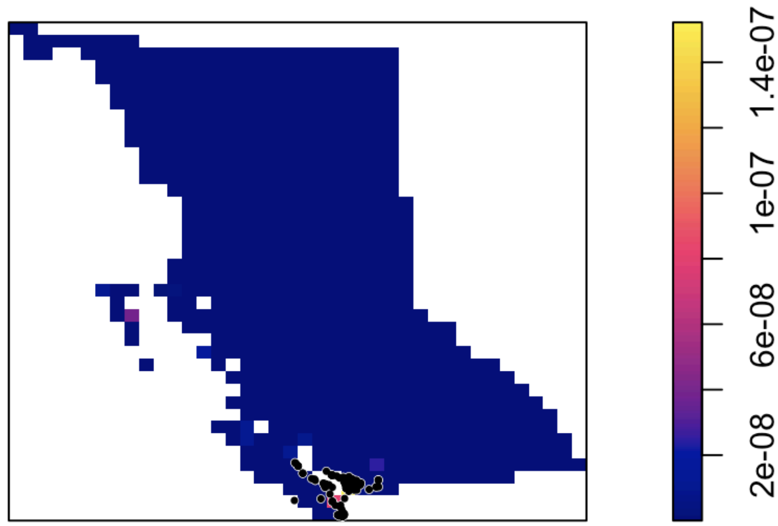

Figure 15. Anna’s Hummingbird predicts Model 10.

Model evaluation

In this section, we evaluate the performance of the selected model number 10, a GAM cubic spline for elevation, forest cover and HFI all with 3 degrees of freedom and distance to water with 8 df (with 4 covariates).

First, a quadrat test conducted on our GAM model indicates a significant deviation from the model's predictions concerning the spatial distribution of Anna's Hummingbirds. This test is valuable as it objectively assesses the agreement between observed and predicted spatial patterns. Although the test identifies a discrepancy, it does not elucidate the underlying cause.

Second, we compute and evaluate the residuals plot. Figure 16 illustrates that our GAM tends to overestimate intensity predictions, particularly near the Pacific coast, where the primary concentrations of hummingbirds are observed in the studied point pattern. This indicates areas where the model performs less accurately.

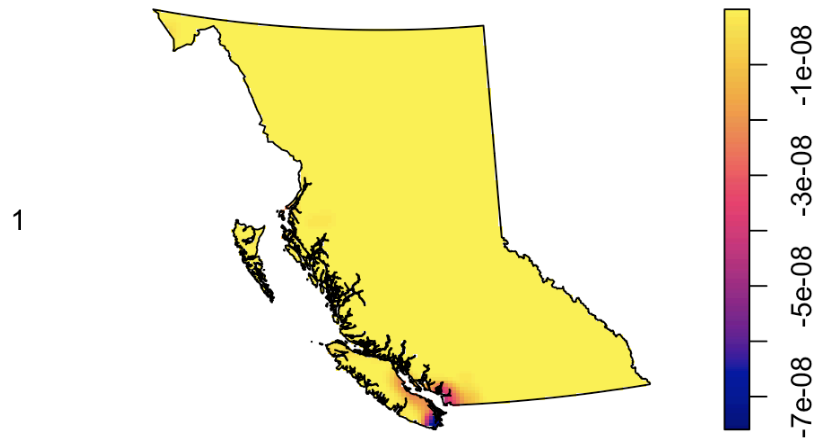

Figure 16. Residuals from the GAM Model 10

As the following step, we plot the partial residuals, which show the fitted effect of a covariate alongside the observed effect. Figure 17 reveals that our complex GAM model struggles to capture elevation patterns between 800 meters and 900 meters. Despite the flexibility inherent in this additive framework model, there are noticeable discrepancies between the predicted and observed values. For instance, the model fails to capture the data patterns above 80% forest cover, as well as at lower Human Footprint Index (HFI) values below 0.2 and for distances to water ranging from 4000 to 8000 meters. The main cause of this phenomenon may be due to the small number of observations which fall in those ranges affecting the prediction accuracy.

Continuing with the analysis of the covariates, with the aim to identify spatial patterns that are not accounted for in the GAM model. We perform lurking variable analysis on each of the predictors. As we can observe in Figure 18 the model is under-estimated when forest cover is 8-10% while it is over-estimated from 10 and 20 percent. This suggests the patterns are not well explained by the selected predictor. For the rest of the covariates the residuals are close to zero and within the confidence intervals, indicating that the model is adequately capturing the spatial pattern present in the data.

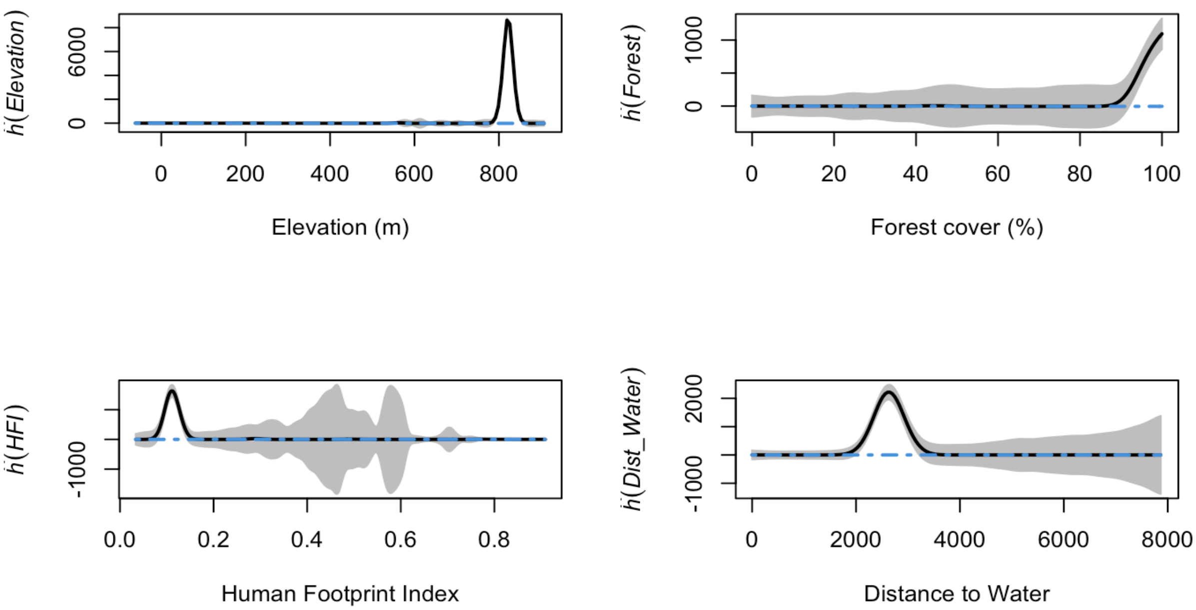

Figure 17. Partial residuals from the GAM Model 10

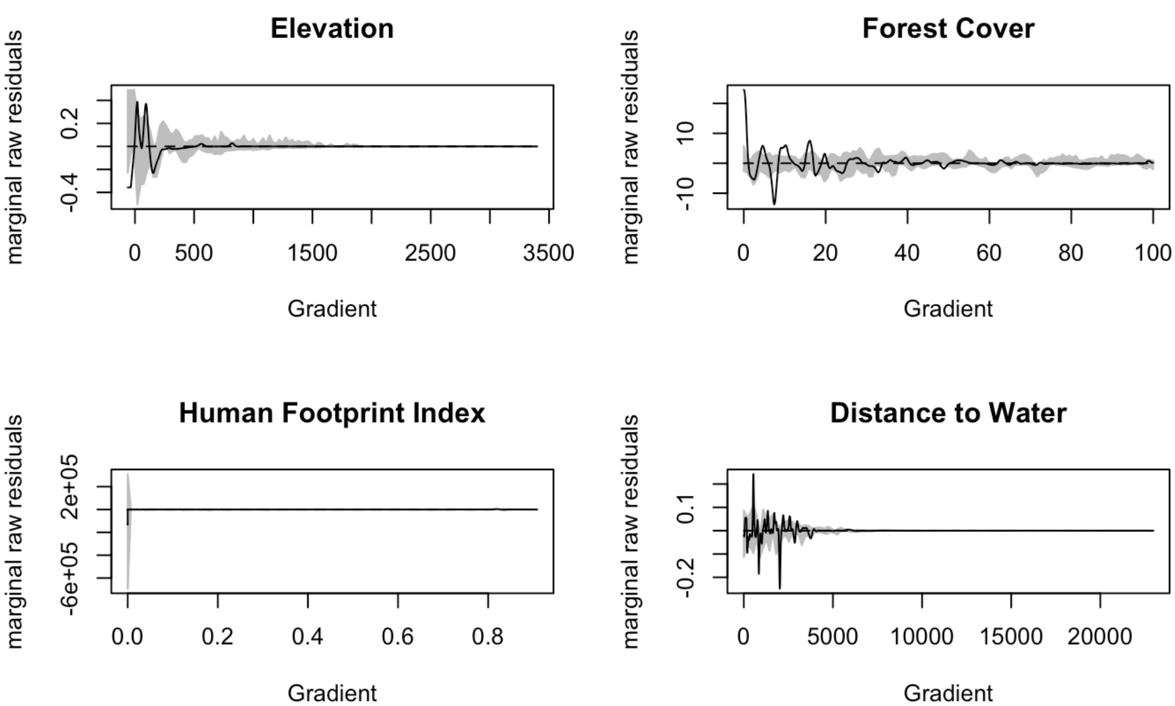

Figure 18. Lurking variables from the GAM Model 10

Discussion

Our analysis uncovered many interesting conclusions about the distribution of Anna's Hummingbirds, and the relative influence each of the covariates have on the data points. Our first hypothesis hinged on our prior knowledge of hummingbirds, which led to a theory that Anna's Hummingbirds will be aligned with human activity (HFI score). We found that there was merit to this claim, specifically the rho estimation for HFI revealed a strongly positive relationship with human activity and hummingbird intensity. However, we found it surprising that Figure 14 had no hummingbird occurrences *at all* in areas with minimal human footprint (0-0.3). Another surprise was the intensity of observations noted in the opposite extreme (HFI values of 0.8-1) were still estimated to have a large concentration of Anna's Hummingbirds. This result is dubious, as intensely urban areas seem unlikely to support any variation of hummingbird. Additionally, this clear positive relationship is muddled, due to our chosen model assigning near-zero coefficients for much of the HFI covariate, leading to conflicting results. One explanation for these results relates to the nature of how this data was collected, which GBIF identifies as 99% human-gathered [3]. As there are less people present in the deep wilderness compared to dense cities, this could explain the unexpected results.

The next hypothesis involves Elevation and Anna's Hummingbirds, and the potential involvement of a preferred temperature, which could serve to motivate this behavior. The Figure 11 rho estimation shows that there is a strong negative correlation, with Anna's Hummingbirds being found almost *exclusively* at low-elevations (0-100m). Supplementary research revealed these hummingbirds are best suited to temperatures between 10-27 degrees Celsius [3], which align with the lowest areas of elevation in BC, where temperature tends to be the highest.

The third hypothesis revolves around the idea of Anna's Hummingbirds having any natural dense forest-dwelling predators, and how that might influence point occurrence in forested areas. Rho estimates from Figure 12 indicate that there is a significant preference amongst hummingbirds for very low (0-20) density forest areas. Research into the topic revealed hawks, owls and snakes are the most common predators of Anna's Hummingbird [7] all of which tend to live in dense forest; lending credence to the idea that these hummingbirds as a whole may avoid those areas.

References

[1] *Anna's hummingbird overview, all about birds, Cornell Lab of Ornithology* (n.d.) *Overview, All About Birds, Cornell Lab of Ornithology*. Retrieved April 27, 2024, from https://www.allaboutbirds.org/guide/annas_hummingbird/#

[2] BirdLife International. 2021. *Calypte anna*. *The IUCN Red List of Threatened Species* 2021: e.T22688199A186913595. Retrieved April 27, 2024, from https://dx.doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22688199A186913595

[3] GBIF.org (13 April 2024) GBIF Occurrence Download https://doi.org/10.15468/dl.zb3e97

[4] Anna's Hummingbird - Calypte anna. (n.d.). Anna's Hummingbird - Calypte anna - Birds of the World. Retrieved April 27, 2024, from https://birdsoftheworld.org/bow/species/annhum/cur/introduction

[5] (n.d.). BC Environment Standard Projection. Retrieved April 28, 2024, from https://ibis.geog.ubc.ca/~brian/Course.Notes/bceprojection.html

[6] British Columbia. (n.d.). Wikipedia. Retrieved April 28, 2024, from https://en.wikipedia.org/wiki/British_Columbia

[7] Atwood, A. (2024, February 23). *Anna's hummingbird*. Hummingbird Facts and Information. https://hummingworlds.com/annas-hummingbird/