586 Project - Oreoscoptes Analysis

ZiYing(Sophie) Chen, Xinyu Dong, Matthew Yau

2023-04-23

Introduction

Wildlife preservation in British Columbia (BC) is an important topic to address. Having diverse wildlife could lead to a number of benefits, such as thriving ecosystems (e.g. diverse species lead to more balanced and resilient ecosystems; Ferrero & Troìa, 2021), or economic benefits (e.g. BC's natural environment is a factor that attracts many tourists yearly, and having wildlife diversity is important to maintaining the health of BC's nature).

One wildlife species that may be important to conserve is the Sage Thrasher (oreoscoptes montanus). In BC, the population of the Sage Thrasher has been in decline, due to a number of factors, such as climate change or human activity (e.g. urbanization or agricultural expansion). For example, estimates have shown up to a 50% decline in Sage Thrasher population over the past century. Furthermore, the Sage Thrasher has also been designated as endangered by BC Ministry of Environment, and is generally considered a rare bird species in Canada (British Columbia Ministry of Environment and Climate Change Strategy, 2019).

As such, the current investigation aims to investigate ways to preserve the conversation of Sage Thrashers in BC by conducting a spatial statistical analysis of the species in BC. In particular, we focus on three questions. First, we examine the spatial distribution of Sage Thrashers in BC (e.g., understanding where Sage Thrashers are situated could help understanding of where to avoid human expansion). Second, we examine the environmental covariates that are linked to the population of Sage Thrashers (e.g., understanding factors that promote or hinder Sage Thrasher population could lead to insights on how to cultivate this species). Third, we build regression models that use covariates to investigate the relationships between geographical intensity of Sage Thrashers and the potential covariates.

Methods

Data Description

To investigate our research question of how we could preserve Sage Thrasher population, we downloaded data of Sage Thrashers from the Global Biodiversity Information Factility (GBIF, 2023). This dataset contains information about where Sage Thrashers have been sighted (e.g., country, coordinates), of which contains 850 samples in British Columbia. To examine how covariates are related to Sage Thrasher populations, we obtained data on BC environmental covariates provided by Michael Noonan (Noonan, 2022). This dataset contains measurements on four environmental covariates (that is, elevation, distance to water, forest, and HFI) across spatial areas in BC (ranging from one million to 7 million measurements).

Detailed Analytical Workflow

Our statistical analysis process involved the following four steps: 1) Preparing the data for analysis, such as obtaining and wrangling data, 2) Conducting spatial distributional analysis of Sage Thrashers, 3) Linking environmental covariates to Sage Thrashers, 4) Creating regression models for Sage Thrashers.

1) Data Preparation

1.1) Load all packages we need. Please refer to **Statistical Packages**.

1.2) Data processing:

- Load Oreoscoptes data by `df <- read.csv("BC_data.csv", stringsAsFactors = FALSE)`.
- Generate a dataframe `coord` with longitude and latitude in BC_data by

  `coord <- df[,c('decimalLatitude','decimalLongitude')]`.
- Clean the data by removing the observations with NA, `coord <- na.omit(coord)`.
- Visualize the data `plot(coord$decimalLatitude, coord$decimalLongitude)`.
- By using the coordinates function to create a new object with the spatial coordinates and then setting its CRS using proj4string, we can ensure that our spatial analyses are performed correctly, and we can compare our results with other analyses that use the same CRS.

```
coordinates(coord)<- ~decimalLongitude+decimalLatitude

proj4string(coord) <- CRS("+proj=longlat +datum=WGS84")

coord_conv <- spTransform(coord,CRS("+proj=aea +lat_0=45 +lon_0=-126 +lat_1=50

              +lat_2=58.5 +x_0=1000000 +y_0=0 +datum=NAD83 +units=m +no_defs"))
```

- Load the covariates data by `load("BC_Covariates.Rda")`.
- Define the observation area, window, by `win = as.owin(DATA$Window)`
- create ppp object with coordination and the window by

```
ppp_birds <- ppp(coord_conv@coords[,1],

             coord_conv@coords[,2],

             window=win)
```

2) Spatial Analysis of Sage Thrashers

2.1) Spatial Analysis of Sage Thrashers

- Check the spread of the bird `plot(ppp_birds)`
- Check homogeneity, perform quardratic test.
  `Q <- quadratcount(bei, nx = 10, ny = 10)`   `quadrat.test(Q)`
- *Hot spot analysis* to identify areas of elevated intensity by

```
## Estimate R
R <- bw.ppl(ppp_birds)

## Calculate test statistic
LR <- scanLRTS(ppp_birds, r = R)

## Plot the output
plot(LR)
```

3) Linking Environmental Covariates to Sage Thrashers

3.1) Linking Environmental Covariates to Sage Thrashers

- Identify the class of each potential covariate by `sapply(DATA, class)`
- Visualise each potential covariate using methods appropriate to each object class.

```
potential <- c("Window", "Elevation", "Forest", "HFI", "Dist_Water")

for (i in potential){
  plot(DATA[[i]], main = i)
  points(ppp_birds$x, ppp_birds$y, pch = 19, col = "white", cex = 0.6)
  points(ppp_birds$x, ppp_birds$y, pch = 19, col = "black", cex = 0.4)
}
```

- Check relationships between potential covariates via *kernel estimation* by `rhohat`

```
rho_elev <- rhohat(ppp_birds, DATA$Elevation)
rho_fore <- rhohat(ppp_birds, DATA$Forest)
rho_hfi <- rhohat(ppp_birds, DATA$HFI)
rho_dist_water <- rhohat(ppp_birds, DATA$Dist_Water)

plot(rho_elev, xlim = c(0, max(DATA$Elevation)))
plot(rho_fore, xlim = c(0, max(DATA$Forest)))
plot(rho_hfi, xlim = c(0, max(DATA$HFI)))
plot(rho_dist_water, xlim = c(0, max(DATA$Dist_Water)))
```

4) Regression Models for Sage Thrashers

4.1) Building the Model

- Scale the covariates by their mean and variance

```
mu_elev <- mean(DATA$Elevation)
stdev_elev <- sd(DATA$Elevation)
DATA$Elevation_scaled <- eval.im((Elevation - mu_elev)/stdev_elev, DATA)

mu_water <- mean(DATA$Dist_Water)
stdev_water <- sd(DATA$Dist_Water)
DATA$Dist_Water_scaled <- eval.im((Dist_Water - mu_water)/stdev_water, DATA)
```

- Establish formula1:
  `formula1 <- ppp_birds ~ Elevation_scaled + I(Elevation_scaled^2) + Dist_Water_scaled + I(Dist_`
- Fit the model with formula1 by `fit1 <- ppm(formula1, data = DATA)`.
- Because the model did not converge, we simplified our formula by excluding *I(Elevation_scaled^2)*.
  `formula2 <- ppp_birds ~ Elevation_scaled + Dist_Water_scaled + I(Dist_Water_scaled^2)`
- Fit the model with formula2 by `fit2 <- ppm(formula2, data = DATA)`.
- Test the statistical significance of coefficients by looking at the Ztest from `fit2` table.
- Identify the formula with coefficients `co <- coef(fit2)`
- Model visualisation

```
plot(fit2,
     se = FALSE,
     superimpose = FALSE)

plot(ppp_birds,
     pch = 16,
     cex = 0.7,
     cols = "white",
     add = TRUE)
plot(ppp_birds,
     pch = 16,
     cex = 0.5,
     cols = "black",
     add = TRUE)
```

- Compute the intensity of a fitted point process model as a function of one of its covariates

```
#Mean disw
#E_disw <- mean(DATA$Dist_Water_scaled) # just 0

#Elevational effect on lambda at mean disw
elev_effect <- effectfun(fit2, "Elevation_scaled", Dist_Water_scaled = 0, se.fit = T)

disw_effect <- effectfun(fit2, "Dist_Water_scaled", Elevation_scaled = 0, se.fit = T)



#Side by side plotting
par(mfrow = c(1,2))

#Plot the elevation effect

plot(elev_effect,
     legend = FALSE,
     main = "Elevation effect at mean distance to water ")
plot(disw_effect,
     legend = FALSE,
     main = "Distance to Water  effect at mean elevation ")
```

4.2) Validating Regression Model

- Quardratic Counting by `quadrat.test(fit2, nx = 2, ny = 4)`
- Create partial residual plots.

```
par_res_elev <- parres(fit2, "Elevation_scaled")

#Calculate the relative intensity as a function of gradient
par_res_disw <- parres(fit2, "Dist_Water_scaled")

#Side by side plotting
par(mfrow = c(1,2))
plot(par_res_elev,
     legend = FALSE,
     lwd = 2,
     main = "",
     xlab = "Elevation_scaled")
plot(par_res_disw,
     legend = FALSE,
     lwd = 2,
     main = "",
     xlab = "Dist_Water_scaled)")
```

4.3) Building Advanced GAM Model

- Build GAM model with 3 degrees of freedom for both covariates.

```
#Fit the PPP model
fit_smooth <- ppm(ppp_birds ~ bs(Elevation_scaled,3) + bs(Dist_Water_scaled, 3), data =
DATA, use.gam = TRUE)
```

- Plot partial residuals for both covariates.

```
#Calculate the partial residuals as a function of elevation
par_res_elev <- parres(fit_smooth, "Elevation_scaled")

#Calculate the relative intensity as a function of gradient
par_res_disw <- parres(fit_smooth, "Dist_Water_scaled")

#Side by side plotting
par(mfrow = c(1,2))
plot(par_res_elev,
     legend = FALSE,
     lwd = 2,
     main = "",
     xlab = "Elevation_scaled")
plot(par_res_disw,
     legend = FALSE,
     lwd = 2,
     main = "",
     xlab = "Dist_Water_scaled)")
```

- Visualize the gam model.

```
#Plot the model predictions
plot(fit_smooth,
     se = FALSE,
     superimpose = FALSE)

#Overlay the B. pendula locations
plot(ppp_birds,
     pch = 16,
     cex = 0.7,
     cols = "white",
     add = TRUE)
plot(ppp_birds,
     pch = 16,
     cex = 0.5,
     cols = "black",
     add = TRUE)
```

Statistical Packages

1. sp: A package providing classes and methods for spatial data: points, lines, polygons and grids
2. dplyr: A Grammar of data manipulation.
3. sf: Create sf object.
4. spatstat: Its main focus is the analysis of spatial patterns of points in two-dimensional space.
5. maptools: Translate and disguise coordinate placing in the real world.
6. rgdal: Provide functions for geospatial data.
7. splines: Create cubic functions.

Results

By looking at Figure 1, we can see that the data is not homogeneous dataset as the birds tend to be clustered in south areas of BC, whereas others have no birds at all, and we conduct a quardratic test to support the inhomogeneous assumption (small p-value). Figure 2 illustrates the hotspot is Osoyoos of the dataset.

Figure 1: Oreoscoptes in BC

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:
## X2 = 71836, df = 63, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 64 tiles (irregular windows)
```

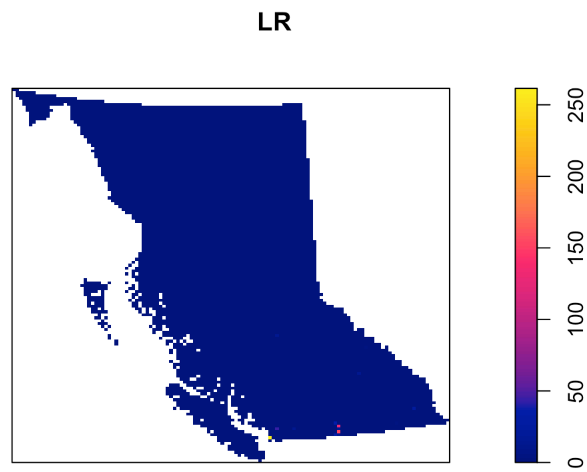

Figure 2: Hotspot of Oreoscoptes in BC

There are four potential covariates in BC_Covariates.Rda, Elevation, Forest, HFI, and Dist_Water. It's hard to tell the relationship between them and intensity, so we used kernel estimation to detect it. Figure 3 shows that there seems to be two clusters: one group living near the 0 elevation, and another group lives in around 500m, and there is a non-linear relationship between elevation and bird intensity. The narrow bandth in Figure 4 describes a specific preference for habitats in terms of the distance to water, and a non-linear relationship between it and bird intensity as well (with one group clustered at 3-5km away from water sources). We exclude `Forest` and `HFI` here because of wider bandwidths.

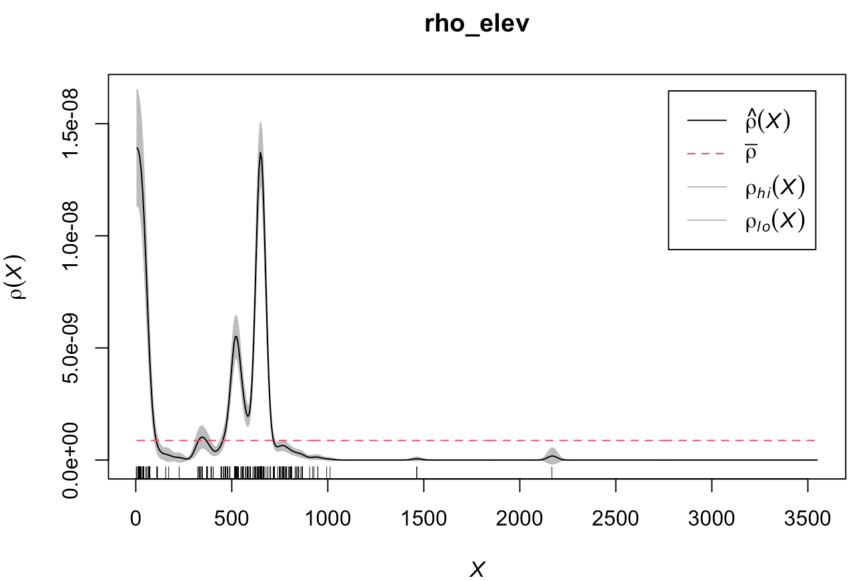

Figure 3: Relationship Diagnosis - Elevation

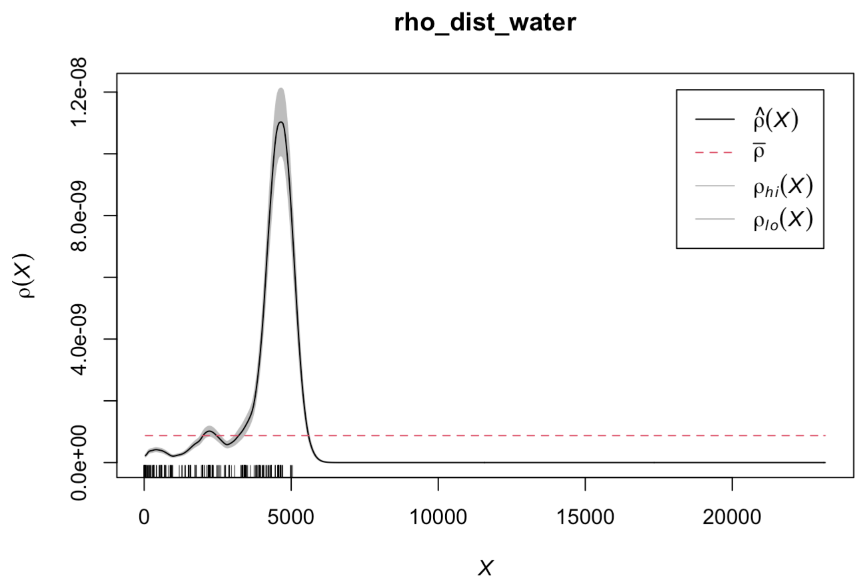

Figure 4: Relationship Diagnosis - Water Distance

After deciding to use `Elevation` and `Dist_Water` to be candidate of covariates with quardratic effect, we scaled both variables to prevent disproportional impact on bird intesity to build the regression model with formula1: ppp_birds ~ Elevation_scaled + I(Elevation_scaled^2) + Dist_Water_scaled + I(Dist_Water_scaled^2).

However, when we tried to fit first model, it failed to converge. Hence, we simplified our formula to formula2: ppp_birds ~ Elevation_scaled + Dist_Water_scaled + I(Dist_Water_scaled^2) by excluding the I(Elevation_scaled^2) and built a second model. By the Ztest from the model, we can say that all the coefficients are statistically significant. It suggested that the intensity can be estimated by the following function:

$\lambda(u)$ = exp(-21.9026746 + -1.6705235 * Elevation_scaled + 1.7131362 * Dist_Water_scaled + -0.549098 * I(Dist_Water_scaled^2))

Figure 5 shows that with these two covariates, the model predicts two novel areas: Vancouver Island and Graham Island as suitable areas to cultivate new Sage Thrasher species.

However, as will be elaborated upon in the discussion section, these suggestions should be interpreted with caution becauase it is possible that our model may not be making these predictions accurately, as we can see that our model performance may not be accurate, as there are no true observations (observation black dots) model's predicted areas (yellow).

To see how individual coefficients affect the intensity, please follow Figure 6. It indicates that `Elevation` has negative marginal effect before it reaches its mean value (`Elevation_scaled` > 0). On the other hand, `Distance_Water` has greatest marginal effect at 2 standard deviation away from its mean value and then the effect goes down.

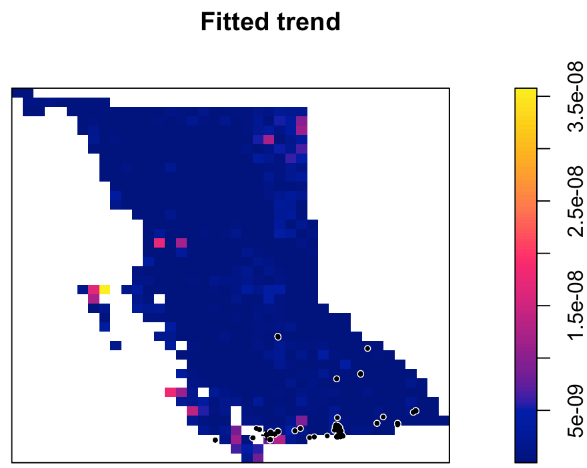

Figure 5: Model Fitting - Fitted Plot

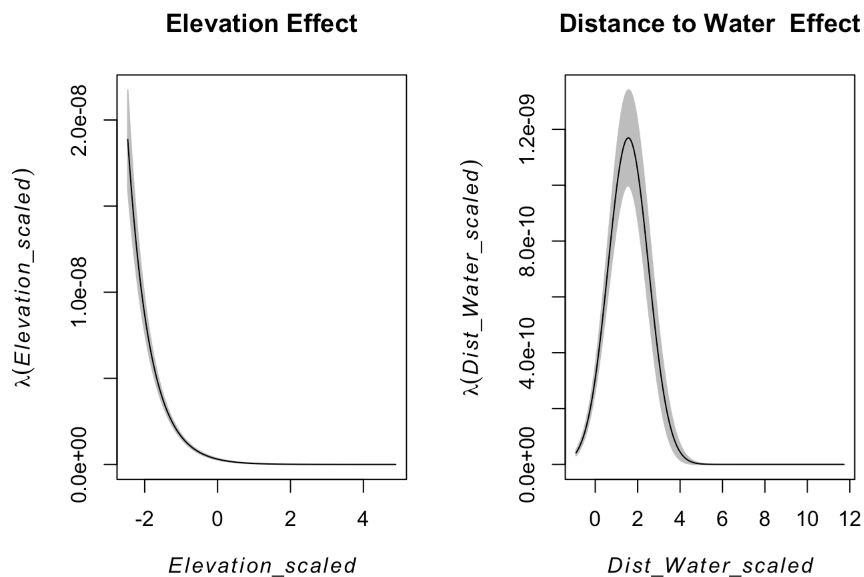

Figure 6: Individual Effect

Though by Figure 5 we can see that the model did not fit well to our dataset, we performed a quadrat test to be more sure about it. The small p-value shows that there's a significant deviation from our model's predictions. Hence, we used the residual plot to see how to enhance it. From Figure 7 we can see that the quadratic terms are not capturing the patterns in our data particularly well. As an improvement, we could try adding higher-order polynomials, but polynomials can be unstable. In this situation, we tried to switch from a linear modelling framework, to an additive modelling framework (i.e., GAMs).

```
##
##  Chi-squared test of fitted Poisson model 'fit2' using quadrat counts
##
## data:  data from fit2
## X2 = 3195.1, df = 4, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 8 tiles (irregular windows)
```

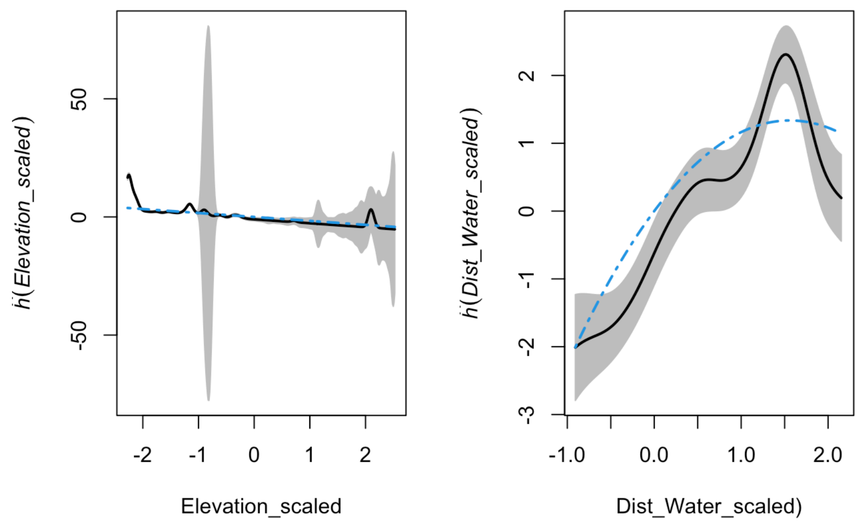

Figure 7: Residual Plot for Model 2

With a gam with degree of freedom 3 for both covariate, Figure 8 shows that the spline term still could not capture the trends. And Figure 9 illustrates that the model still not fit the dataset.

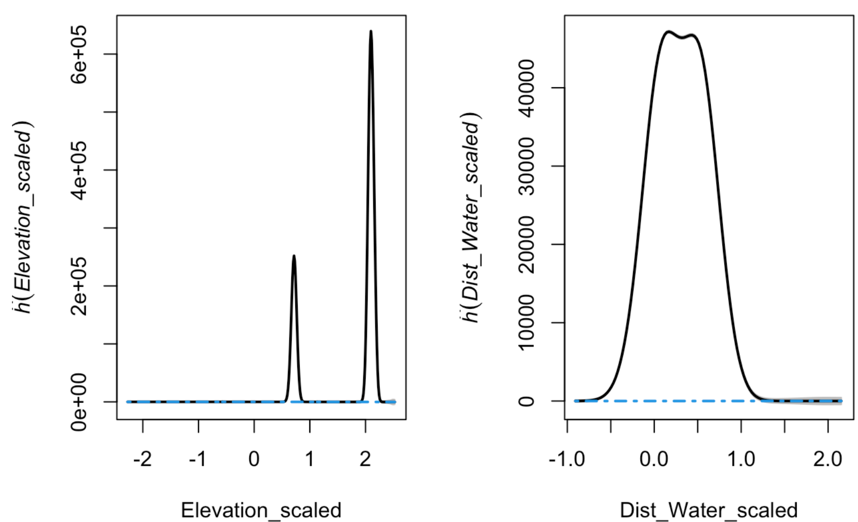

Figure 8: Residual Plot for GAM

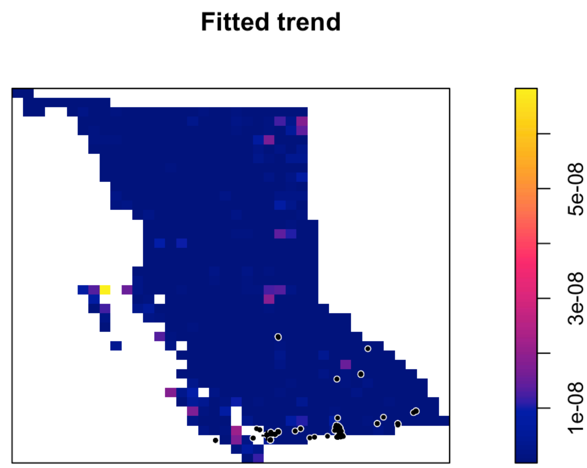

Figure 9: Model Fitting - GAM Fitted Plot

Discussion

The present investigation aimed to examine how Sage Thrasher population could be preserved in BC by focusing on three questions. First, we investigated the spatial distribution of Sage Thrashers in BC. Our analysis showed that Sage Thrashers are significantly disproportionately located in the southern areas of BC. Second, we investigated covariates that are linked to Sage Thrasher populations. Our analysis found that certain elevational height (i.e., low elevations of 0 and 500 meters) and distance to water (i.e. proximities of 3000-5000 meters to water) are covariates that are positively linked to sage thrasher intensity (e.g., populations) in BC. Third, we performed regression model to predict sage thrasher spatial intensity given the potential covariates. Our model predicted, given covariates that have been positively linked to sage thrasher intensity in the past, the novel areas of Vancouver Island and Graham Island as two new areas that contain high intensities of sage thrasher populations.

The findings from the present investigation provide insights into ways in which sage thrashers could be conserved in BC. First, it suggests that, areas in southern BC (where sage thrashers are disproportionately located) should be avoided for industrial development. Second, it suggests that the creation of habitats with elevations of 0 or 500 meters, and 3 to 5 kilometers away from water sources could be beneficial for sage thrasher populations. Lastly, it suggests that Vancouver and Graham Island are potential favorable areas, with similar environmental conditions that have been previously linked to healthy sage thrasher populations (i.e., distance to water source of 3-5km, elevation of 0 to 500 meters). These new areas with favorable environmental conditions could be taken into consideration for the cultivating new populations for sage thrashers.

Although the current analysis provides several insights into Sage Thrasher conservation, there are also a number of limitations. One limitation relates to the statistical validity of our regression models. For example, although our regression model predicted areas in which we could cultivate species of sage thrashers (e.g., Vancouver Island), it should be noted that we also observed that our model did not fit the dataset well. Residual plots revealed that the quadratic terms were not capturing the patterns in the data particularly well. Furthermore, advanced modelling techniques by fitting the data using a GAM with degree of freedom of 3 for both covariates, also failed to fully trends in the data effectively. Thus, future work should aim to improve upon analytical spatial regression modelling techniques to better predict sage thrasher spatial intensity.

A second limitation relates to the methodology of the study. For example, given the rarity of Sage Thrashers in BC, there was only a small sample size of Sage Thrashers (i.e. around 800 Sage Thrashers) to base our analysis upon, which could limit the statistical accuracy of our findings (e.g., higher likelihood of statistical errors due to small samples). Second, the current analysis was only able to obtain a small number of environmental covariates to link Sage Thrasher populations to, which may hinder the ability to understand true and more important predictors that may allow Sage Thrasher populations to thrive (e.g., weather). As such, future studies should aim to incorporate higher sample sizes of Sage Thrashers, and greater numbers of environmental covariates to link to Sage Thrashers.

References

British Columbia Ministry of Environment and Climate Change Strategy. (2019). Sage Thrasher. Government of British Columbia. https://www2.gov.bc.ca/assets/gov/environment/plants-animals-and-ecosystems/species-ecosystems-at-risk/brochures/sage_thrasher.pdf (https://www2.gov.bc.ca/assets/gov/environment/plants-animals-and-ecosystems/species-ecosystems-at-risk/brochures/sage_thrasher.pdf)

Ferrero, P., & Troìa, A. (2021). The Role of Biodiversity in Ecosystem Resilience. Sustainability, 13(14), 7681. https://doi.org/10.3390/su13147681 (https://doi.org/10.3390/su13147681)

GBIF.org. (2023). Ixodes ricinus Linnaeus, 1758. GBIF Occurrence Download. https://www.gbif.org/species/2494973 (https://www.gbif.org/species/2494973)

Noonan, M. (2022, January 15). Data-589 [GitHub repository]. GitHub. https://github.com/ubco-mds-2022/Data-589 (https://github.com/ubco-mds-2022/Data-589)