# Red Foxes in BC: an Analysis of the Spatial Point Process

Nowshaba Durrani, Ricky Heinrich, Viji Rajagopalan

2023-04-09

## Introduction

For our Data 589 project, we have selected Red Fox (Scientific Name - Vulpes Vulpes) to do the analysis. In the GBIF database they have approximately, 610,958+ georeferences records for this species around the world, however for this project we have selected to do the analysis of the occurance of Red Fox in BC only. So with the above function we have fetched the information for British Columbia only in 127 columns and 242 number of entries.

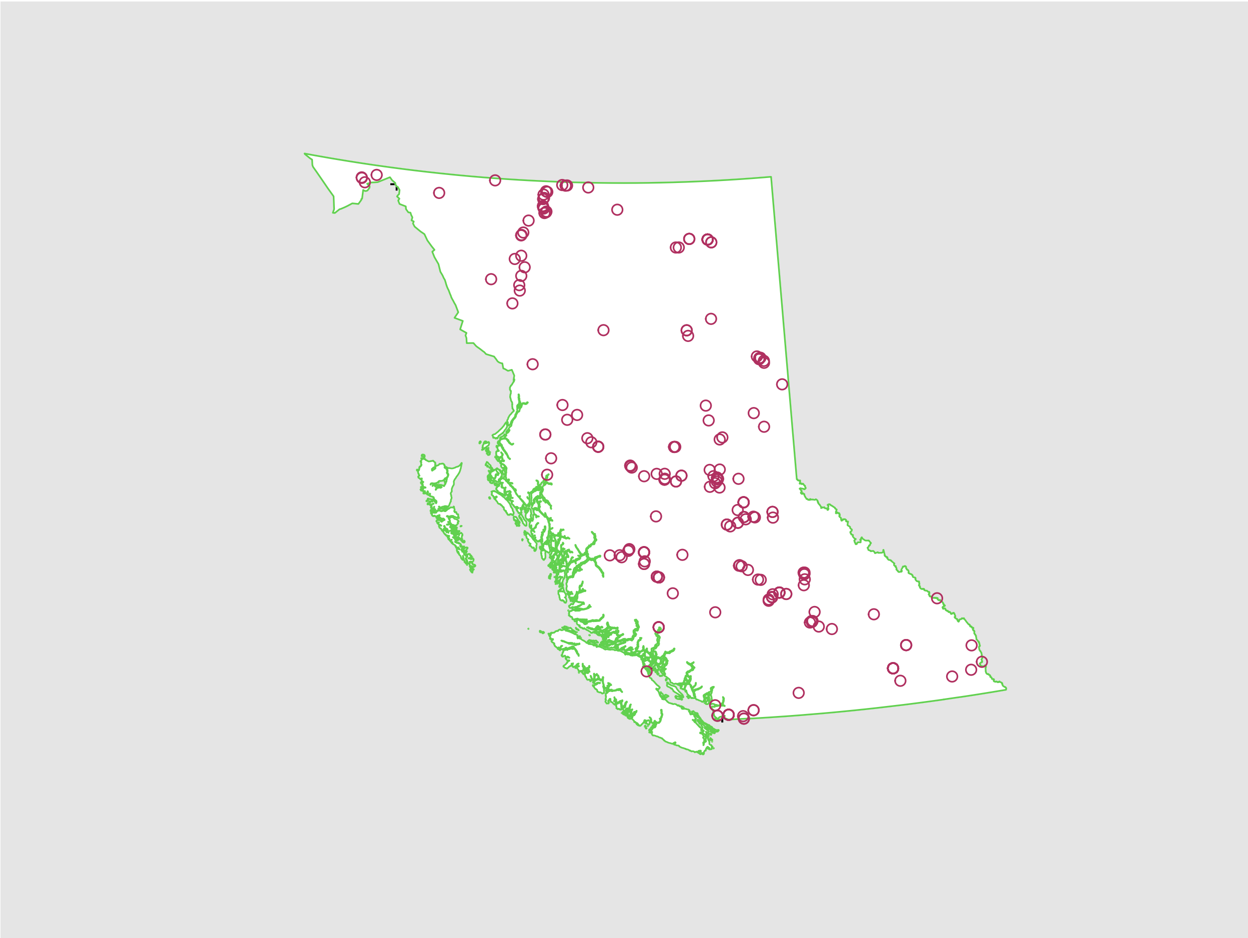


Figure 1: Occurence of Red Foxes in BC

Here we have plotted all the occurrences of Red Fox in the BC region and we can see that the species are scattered in the region specially in the upper and middle part of the province. Now we will be exploring what is contributing to the occurrences of the species in the specific places based on various factors like elevation, close to water bodies, forests, human habitats, etc.

# Methods

Briefly describe the data and what variables are included. Provide a detailed description of the analytical workflow that was applied to the data, citing any relevant literature and statistical packages employed. There should be enough information that anyone can reproduce the workflow if they had access to the data. Length: As long as necessary.

The data comes from the Global Biodiversity Information Facility (GBIF) databases. We used the package `rgbif` to access the 'Vulpes Vulpes' data from R directly, sorting by instances occuring in BC. We've extracted the longitude and latitude data from this, and converted it appropriately using the `sp` package.

Our second source of data contained the BC Window object, as well as possible covariate data: elevation, forest cover, Human Footprint Index (HFI) and distance to water.

We used the package `spatstat` to build a `ppp` object with the converted coordinates of the Red Fox locations from the GBIF data and the window from our second data source.

To conduct first moment analysis, we used functions from the aforementioned `spatstat` package. We did a quadrat test as well as hotspot analysis to gain insight into the homogeneity assumption of the point process.

For second moment analysis, we looked into Ripley's K-function and pair correlation function using functions from `spatstat`. This provides us with insight into possible clustering tendencies of the point process.

Next we looked into the relationship of the intensity with each covariate.

# Results

## Exploratory Analysis (First Moment till Covariates & basic individual models)

## First Moment Analysis

We start with investigating whether the occurence of red foxes in BC seems homogeneous, as it will inform our steps to define the intensity. We have conducted a quadrat test of homogeneity with both 5 x 5 and 10 x 10 quadrats. These quadrats are shown in Figure 2, where we can visually tell that the intensity in each quadrats are not the same. The quadrat test for both the 5 x 5 and the 10 x 10 quadrats provide a p-value of 2.2e-16, confirming that the varied intensities are not due to chance alone, but rather due to an inhomogeneous point process.

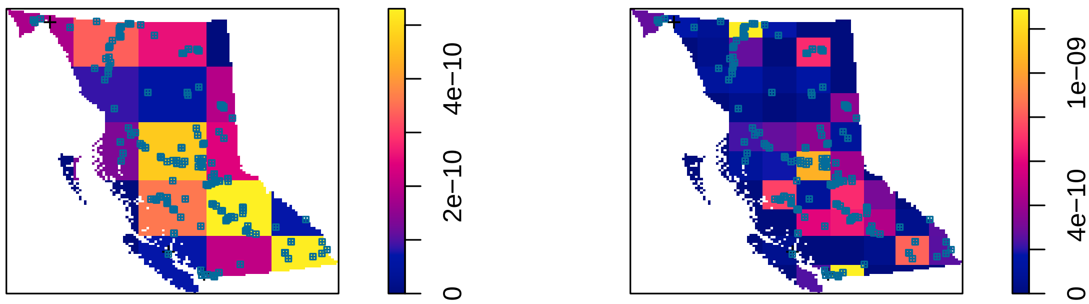



Figure 2: Intensity of Quadrat counts of Red Fox occurences, left 5x5, right 10x10

As the next step, we investigate analyze for any hot spots in the occurences of red foxes. In Figure 3, we can see that hotspots appear scattered around the province. It seems like the highest densities of red fox occurences are more inland, but spread across the province.

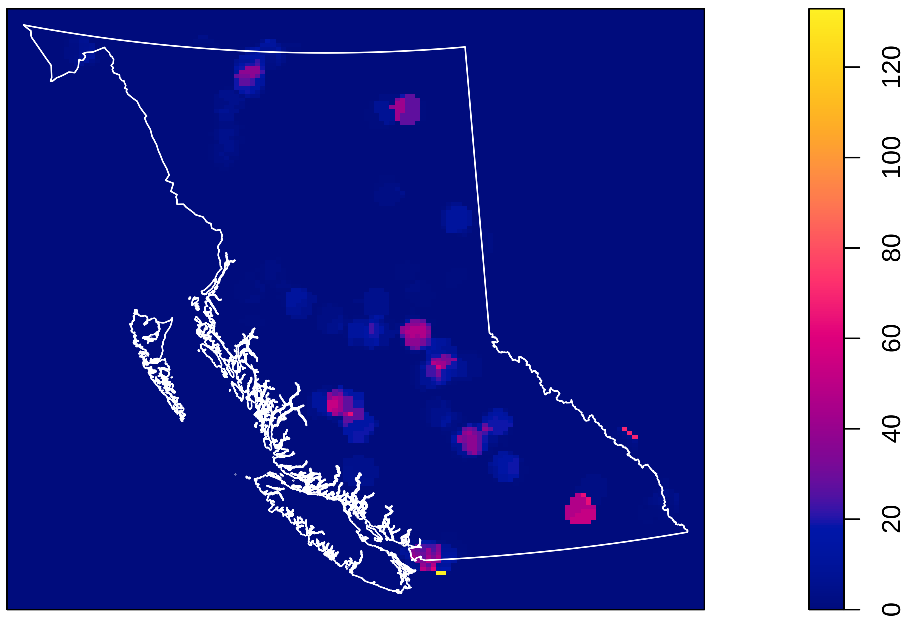



Figure 3: Hotspot of Red Foxes

## 2nd Moment Analysis

Ripley's K-function provides information on whether there are significant deviations from independence between points. Taking into account that the intensity of red fox occurences appear inhomogeneous, we can see in Figure 4 that there is some evidence of clustering up to a certain distance, as the black line, indicating the observed data, is separate from the 95% confidence bands of the values expected with no clustering. This suggests that the relationship between points may be due to effects between points rather than relationship with covariates. [CHECK THIS CLAIM ?]

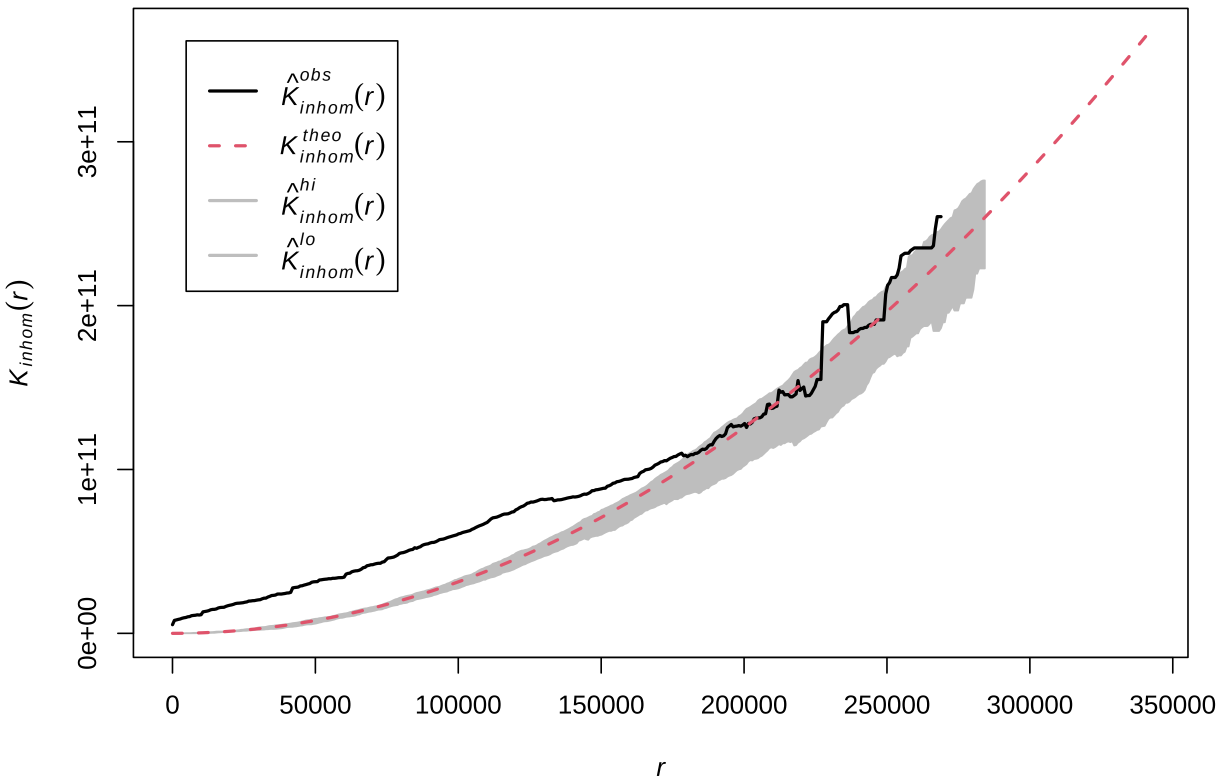



Figure 4: Ripley's K function with border correction assuming inhomogeneity

To get a sense of the distances for which clustering occurs, we used the pair correlation function. Figure 5 shows evidence for clustering at distances smaller than around 23 000m, or 23km but after that the observed values are not significantly different that those expected from a random spatial process.

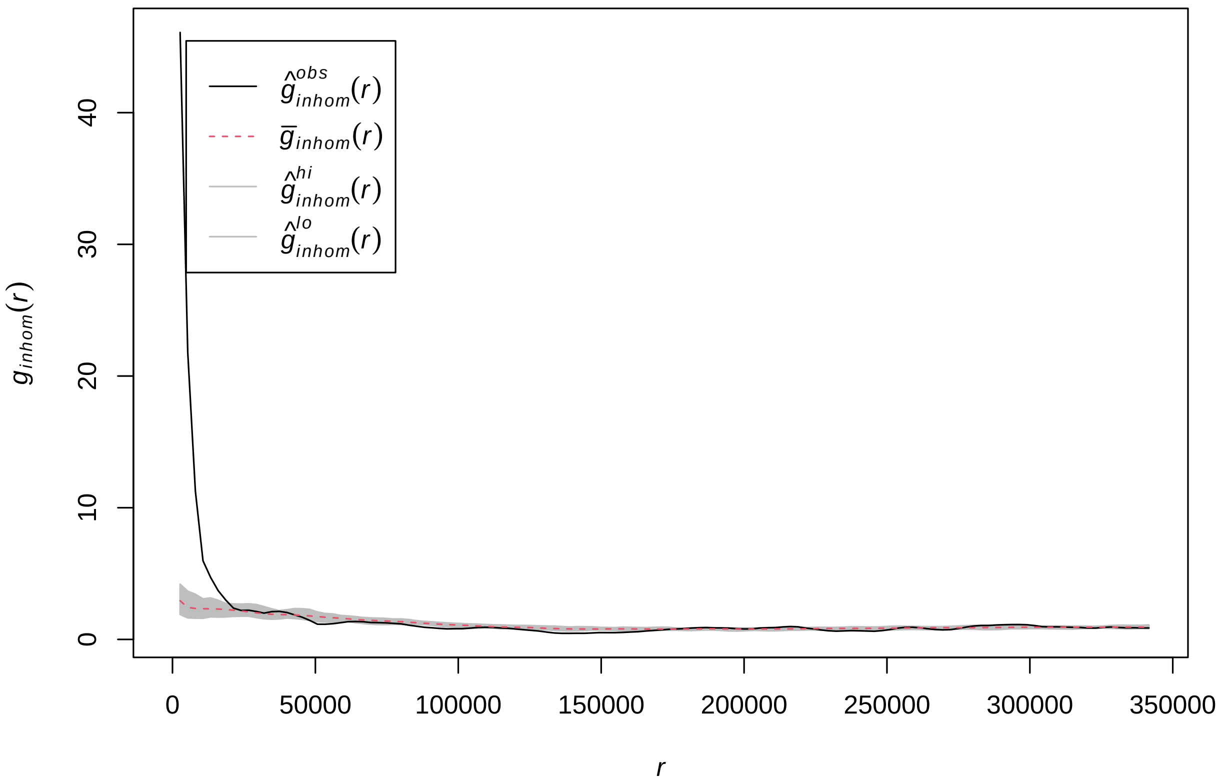



Figure 5: Pair correlation function assuming inhomogeneity

## Relationship with Covariates

Our data includes 4 covariates which we are exploring: the elevation, the forest cover, the human footprint inventory (HFI), and the distance to water. Given our research questions, we were expecting HFI and forest cover to have a relationship, but we have also investigated the two other covariates.

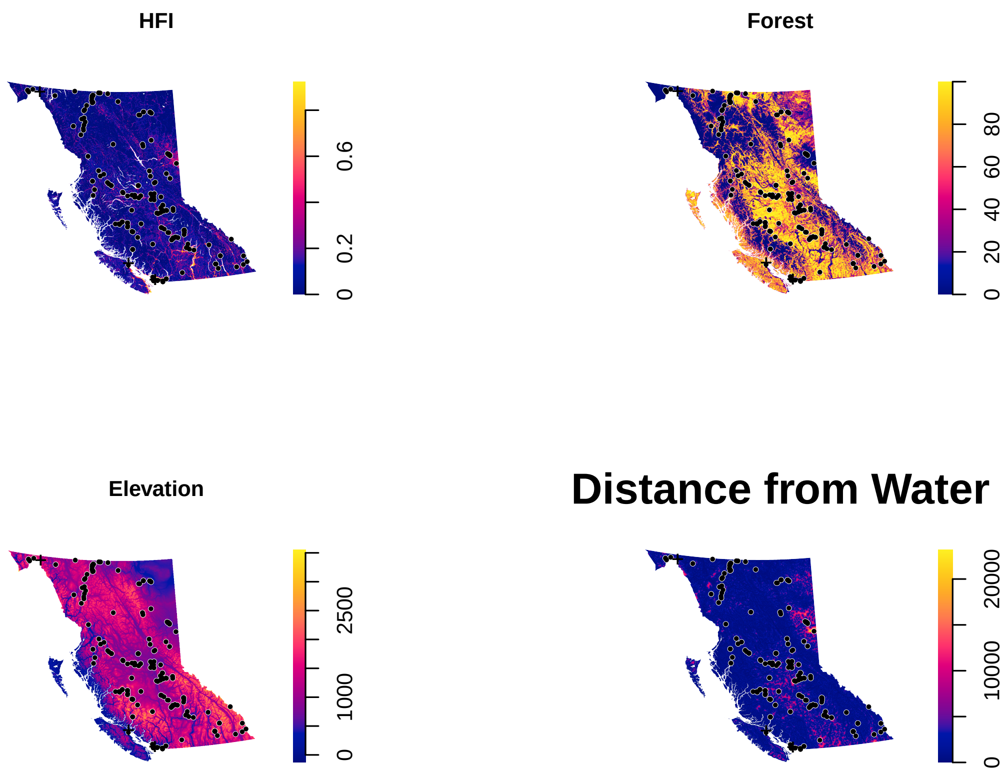



Figure 6: Red Fox plotting in various covariets

In Figure 6, we can see the observed data over the values of the covariates, and get a first idea of possible relationships. We see the red fox is seen more in the averagely densed population area and near highly densed forest area. In case of elevation, red foxes are seen in moderately elevated area and where the distance to water is low to medium.

**Smoothing Estimate of the 4 Covariates Transformation**

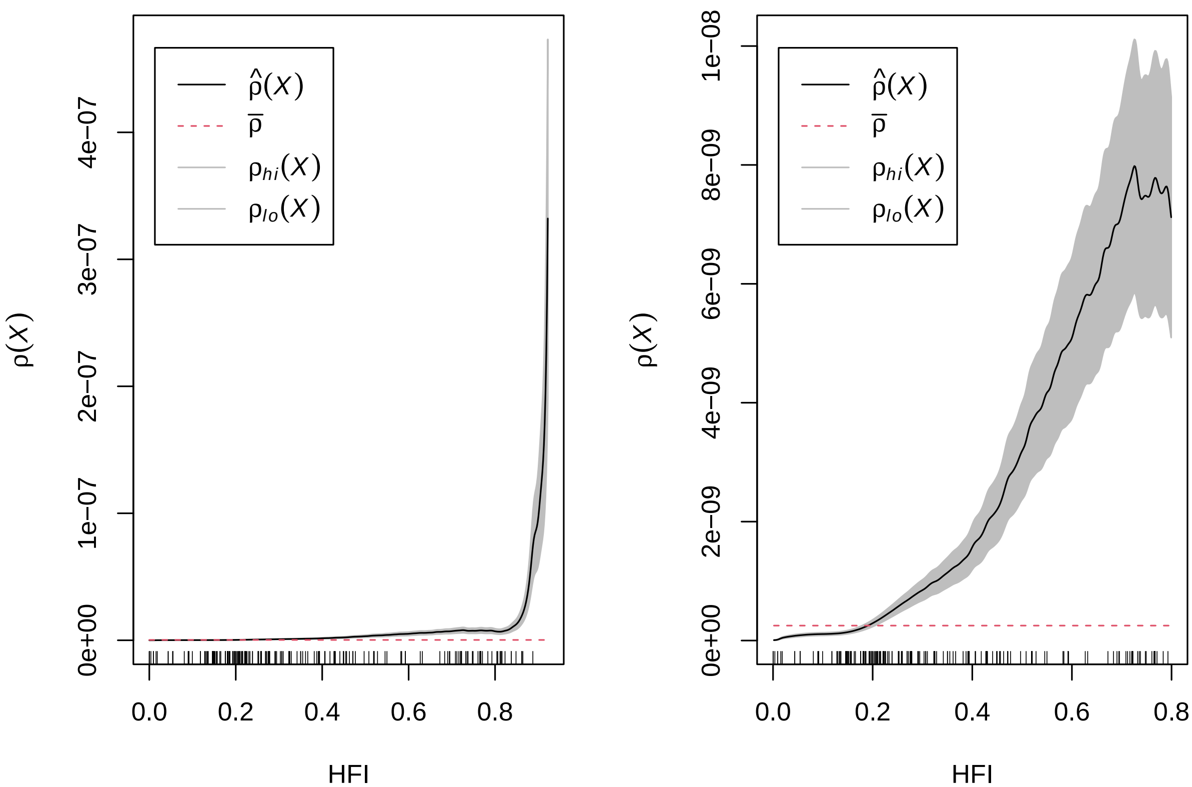



Figure 7: Effect of HFI on intensity of red foxes

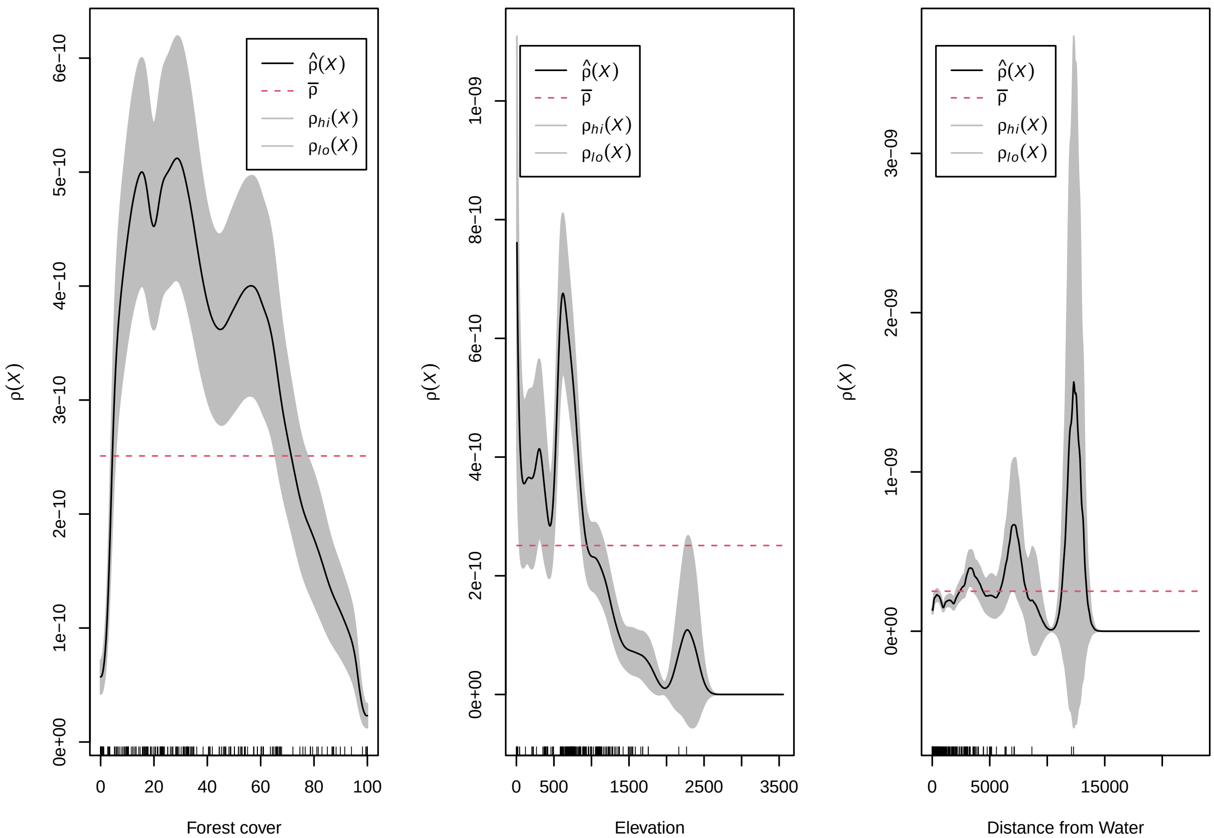



Figure 8: Effect of forest cover, elevation, and distance to water on intensity of red foxes

In the left plot of Figure 7, we could be fooled into thinking that there is no relationship between HFI and intensity of red foxes up to around HFI = 0.4, until which it seems like an exponential relationship. However, zooming in from HFI 0 to 0.8, we see that the confidence bands don't intersect at all with the red line, which is the expected value given no relationship. This relationship appears non-linear and possibly exponential, where the greatest intensity of observed red foxes occurs at high HFIs. This relationship was expected, as our dataset is not exhaustive but rather is crowdsourced, and naturally foxes are more likely to be observed by humans in spaces with higher HFIs.

In Figure 8, we see that there seems to be non-linear relationship between forest cover and number of red foxes observed. The observance increase with increase in forest cover at intermediate coverage and then it decreases. We also see that there is non-linear relationship with elevation. The relationship appears to be non-linear as the graph is showing different results for different elevation and we cannot see any type of specific pattern from the same. In case of Distance to Water, we don't observe a significant deviation in observed red foxes than expected by chance, indicating that it is not a useful covariate to model.

**Fit models for the covariets**

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~HFI
```

```
##
## Fitted trend coefficients:
## (Intercept)         HFI
##   -23.31421     5.98177
##
##              Estimate      S.E.    CI95.lo    CI95.hi Ztest       Zval
## (Intercept) -23.31421 0.1058336 -23.521645 -23.106785   *** -220.29132
## HFI           5.98177 0.2148471   5.560677   6.402862   ***   27.84199
## Problem:
##  Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~HFI + exp(HFI)
##
## Fitted trend coefficients:
## (Intercept)         HFI    exp(HFI)
##   -13.71641    22.17067   -10.35893
##
##              Estimate     S.E.   CI95.lo    CI95.hi Ztest      Zval
## (Intercept) -13.71641 1.404309 -16.46881 -10.964015   *** -9.767372
## HFI          22.17067 2.382441  17.50117  26.840165   ***  9.305861
## exp(HFI)    -10.35893 1.522442 -13.34286  -7.374997   *** -6.804153
## Problem:
##  Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~Forest + I(Forest^2)
##
## Fitted trend coefficients:
##   (Intercept)        Forest   I(Forest^2)
## -2.225373e+01  3.996522e-02 -5.280288e-04
##
##                  Estimate         S.E.       CI95.lo       CI95.hi Ztest
## (Intercept) -2.225373e+01 1.330469e-01 -2.251450e+01 -2.199297e+01   ***
## Forest       3.996522e-02 7.091557e-03  2.606603e-02  5.386442e-02   ***
## I(Forest^2) -5.280288e-04 7.699618e-05 -6.789385e-04 -3.771191e-04   ***
##                   Zval
## (Intercept) -167.262353
## Forest         5.635606
## I(Forest^2)   -6.857857
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~Elevation
##
## Fitted trend coefficients:
```

```
##   (Intercept)     Elevation
## -20.775522387  -0.001401046
##
##                  Estimate         S.E.       CI95.lo       CI95.hi Ztest
## (Intercept) -20.775522387 0.1298745141 -21.030071757 -20.520973016   ***
## Elevation    -0.001401046 0.0001421902  -0.001679734  -0.001122358   ***
##                    Zval
## (Intercept) -159.966122
## Elevation     -9.853321
```

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~Dist_Water
##
## Fitted trend coefficients:
##   (Intercept)    Dist_Water
## -2.211986e+01  1.071969e-05
##
##                  Estimate         S.E.       CI95.lo       CI95.hi Ztest
## (Intercept) -2.211986e+01 8.745074e-02 -2.229126e+01 -2.194846e+01   ***
## Dist_Water   1.071969e-05 3.419673e-05 -5.630466e-05  7.774404e-05
##                     Zval
## (Intercept) -252.9408319
## Dist_Water     0.3134712
```

We have fitted 6 models and we came to observe that HFI, exp(HFI), Forest, I(Forest^2) and elevation seems to be highly significant however Distance to Water seems to be in-significant for the occurrence of red foxes in the BC area.

```
## [1] 10468.83
```

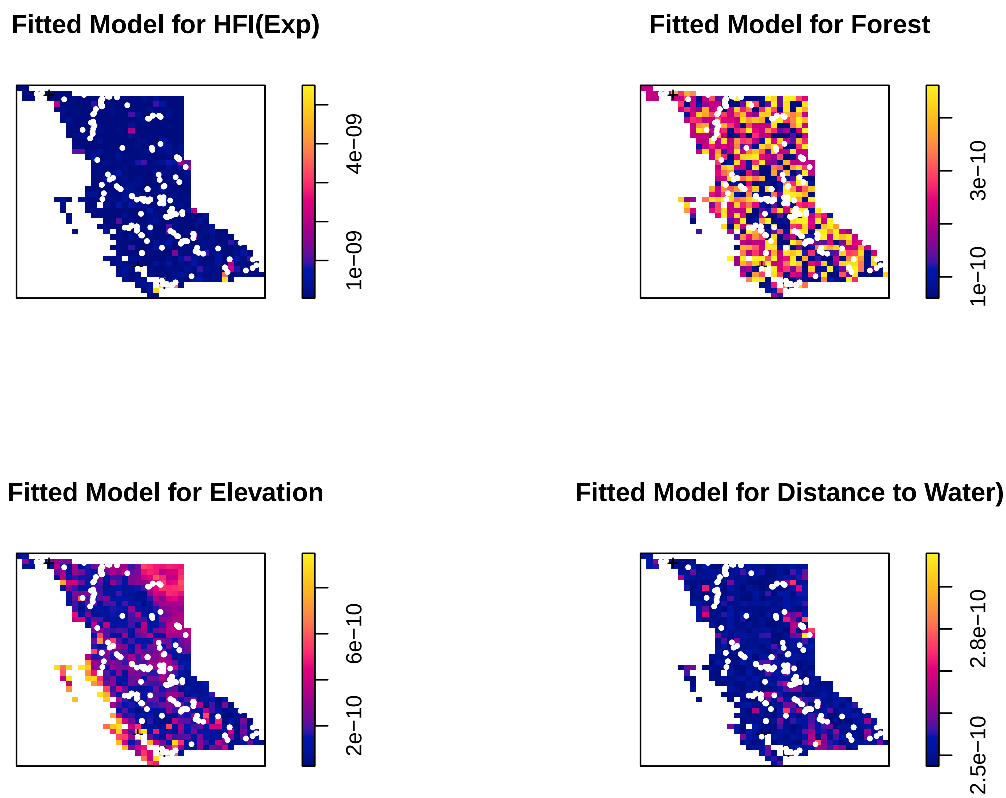
```
## [1] 10420.78
```

```
## [1] 10931.26
```

```
## [1] 10896.99
```

```
## [1] 11000.3
```

## Plot the fitted models




Does not look like a good fit for all points in forest and also not easy to interpret. Here Elevation is significant but Dist to Water is not significant.

### 2. Checking collinearity

```
##              ..1         ..2         ..3         ..4
## ..1  1.00000000  0.06616335 -0.26217406  0.04822162
## ..2  0.06616335  1.00000000 -0.26625709  0.13249159
## ..3 -0.26217406 -0.26625709  1.00000000 -0.03497584
## ..4  0.04822162  0.13249159 -0.03497584  1.00000000
```

We see that no covariate is strongly collinear with another, and so we can move on without taking so into account and treating the covariates as independent.

## End of EDA and covariates analysis

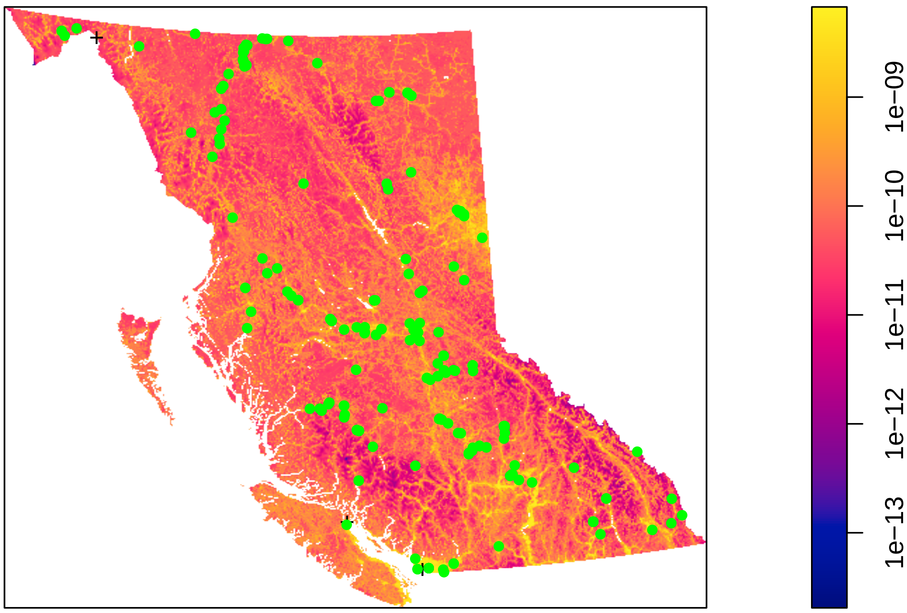
## Model Fitting

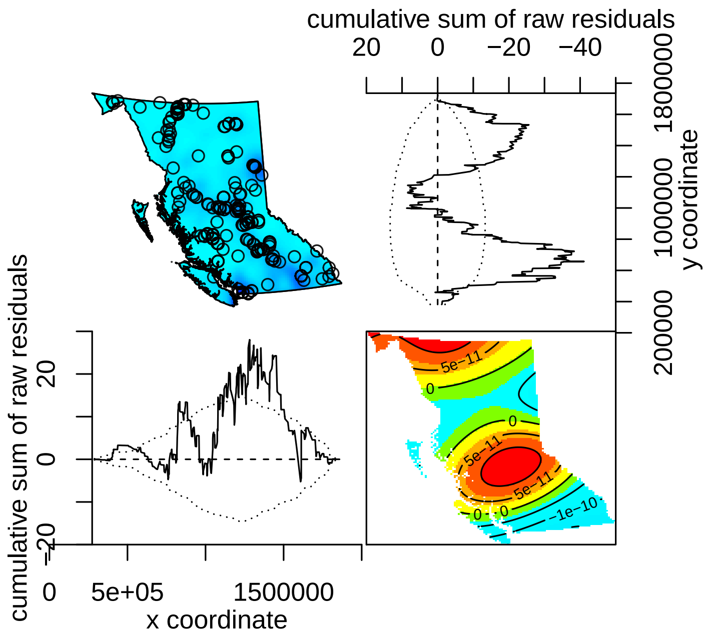
Try different models - linear and quadratic, do necessary anova and other tests for model selection and validation

```
## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~Forest + I(Forest^2) + HFI + I(HFI^2) + Elevation +
## I(Elevation^2)
##
## Fitted trend coefficients:
##    (Intercept)         Forest    I(Forest^2)            HFI       I(HFI^2)
##  -2.413811e+01   6.206401e-03  -1.160232e-04   1.185642e+01  -7.371997e+00
##      Elevation I(Elevation^2)
##   1.364005e-03  -9.678328e-07
##
##                     Estimate         S.E.       CI95.lo       CI95.hi Ztest
## (Intercept)    -2.413811e+01 3.220178e-01 -2.476925e+01 -2.350696e+01   ***
## Forest          6.206401e-03 7.444014e-03 -8.383599e-03  2.079640e-02
## I(Forest^2)    -1.160232e-04 7.767304e-05 -2.682596e-04  3.621315e-05
## HFI             1.185642e+01 1.088713e+00  9.722583e+00  1.399026e+01   ***
## I(HFI^2)       -7.371997e+00 1.232831e+00 -9.788301e+00 -4.955694e+00   ***
## Elevation       1.364005e-03 5.226588e-04  3.396126e-04  2.388397e-03    **
## I(Elevation^2) -9.678328e-07 2.819525e-07 -1.520449e-06 -4.152160e-07   ***
##                       Zval
## (Intercept)    -74.9589241
## Forest           0.8337439
## I(Forest^2)     -1.4937385
## HFI             10.8903097
## I(HFI^2)        -5.9797320
## Elevation        2.6097427
## I(Elevation^2)  -3.4326094
## Problem:
##  Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points
```

Forest related variables are not highly significant when compared to HFI and elevation but still significant. So we retain them in the model.

## Fitted trend

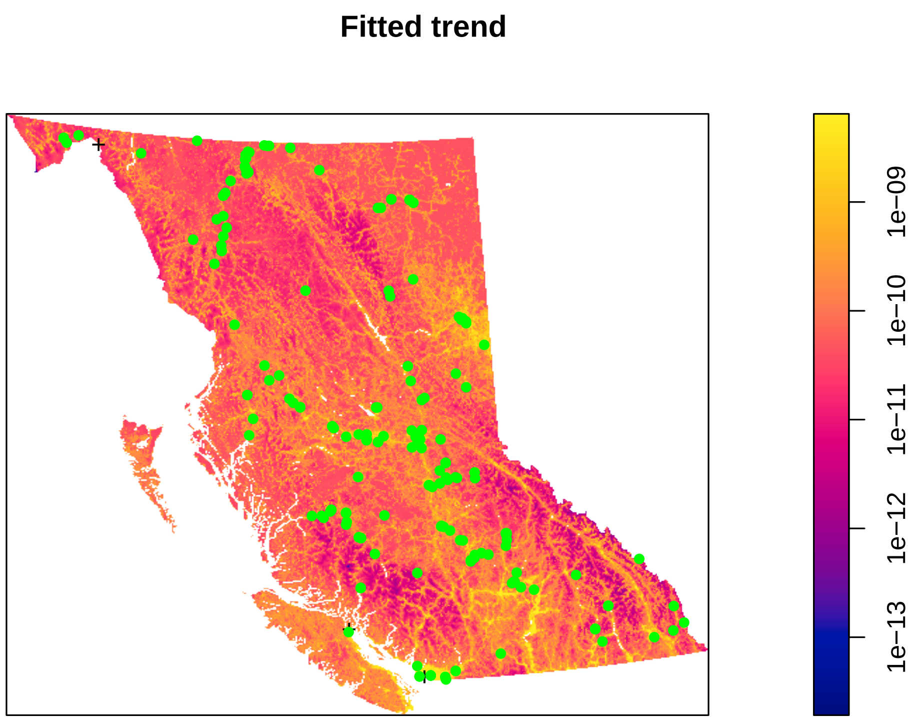





```
## Model diagnostics (raw residuals)
## Diagnostics available:
##  four-panel plot
##  mark plot
##  smoothed residual field
##  x cumulative residuals
##  y cumulative residuals
##  sum of all residuals
## sum of raw residuals in entire window = -4.089e-06
## area of entire window = 9.483e+11
## quadrature area = 9.39e+11
## range of smoothed field =  [-1.518e-10, 1.315e-10]

## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~HFI + I(HFI^2) + Elevation + I(Elevation^2)
##
## Fitted trend coefficients:
##    (Intercept)            HFI        I(HFI^2)      Elevation I(Elevation^2)
##  -2.426748e+01   1.242099e+01  -7.734861e+00   1.255468e-03  -9.320293e-07
##
##                        Estimate         S.E.       CI95.lo       CI95.hi Ztest
```

```
## (Intercept)    -2.426748e+01 2.983606e-01 -2.485226e+01 -2.368270e+01   ***
## HFI             1.242099e+01 1.073719e+00  1.031654e+01  1.452544e+01   ***
## I(HFI^2)       -7.734861e+00 1.222941e+00 -1.013178e+01 -5.337941e+00   ***
## Elevation       1.255468e-03 5.201189e-04  2.360535e-04  2.274882e-03     *
## I(Elevation^2) -9.320293e-07 2.837978e-07 -1.488263e-06 -3.757958e-07    **
##                      Zval
## (Intercept)    -81.336078
## HFI             11.568195
## I(HFI^2)        -6.324804
## Elevation        2.413809
## I(Elevation^2)  -3.284131
## Problem:
##  Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points
##
## *** Fitting algorithm for 'glm' did not converge ***
```




```
## [1] 10401.62
```

```
## [1] 10403.57
```

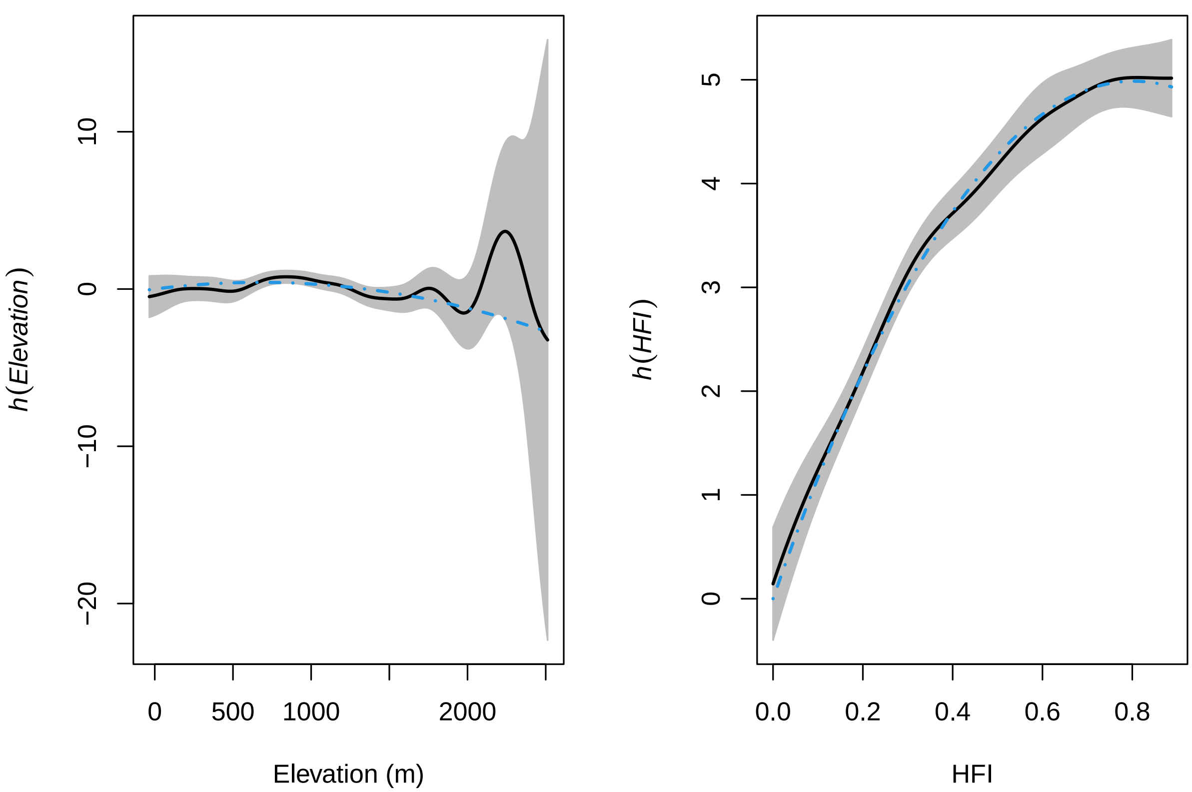
**Validate if we need GAM - 5. linear vs gam - table: evaluate, AIC, visualize**

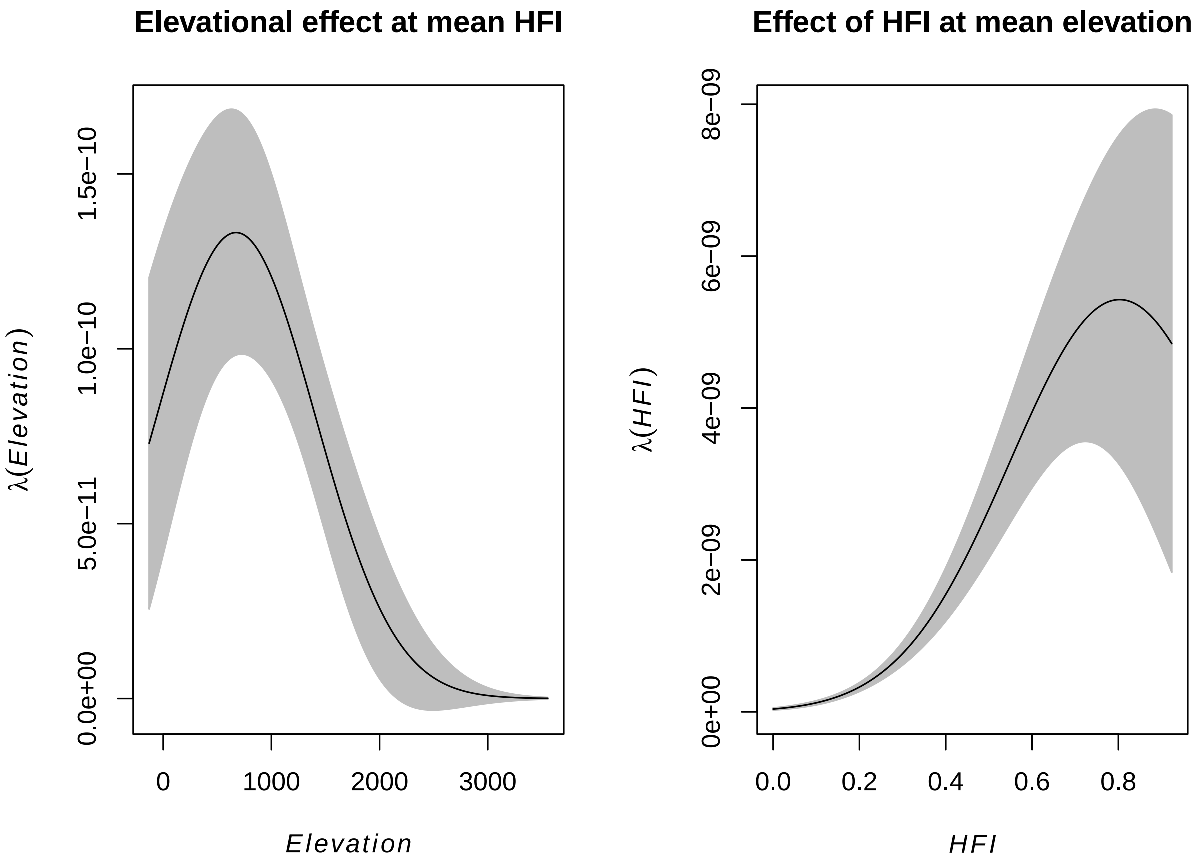
```
##
##  Chi-squared test of fitted Poisson model 'fit_red2' using quadrat
##  counts
##
## data:  data from fit_red2
## X2 = 152.84, df = 16, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 21 tiles (irregular windows)
```

Rho plots do not run due to NA error.

# using above, check the intensity.

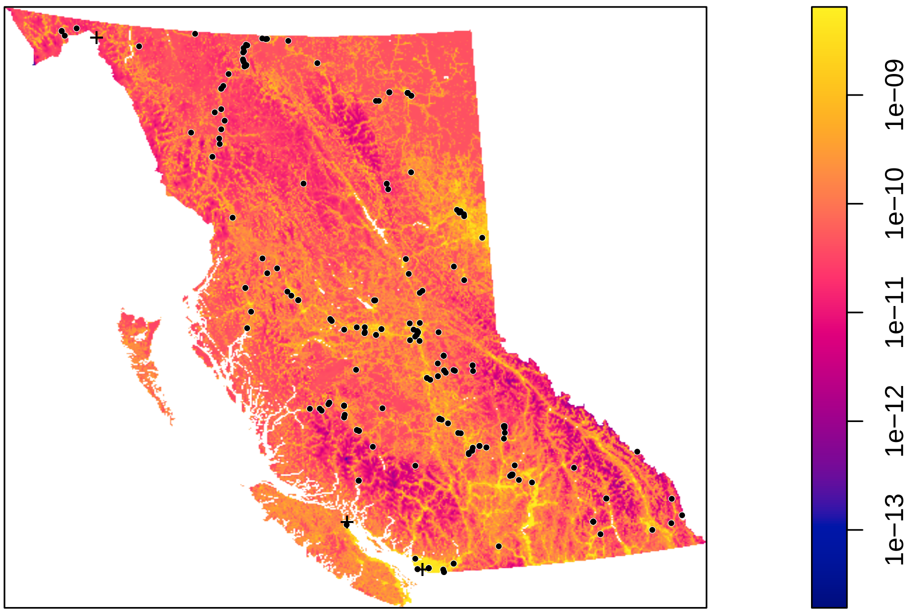
##Calculate partial residuals only for 2 covariates

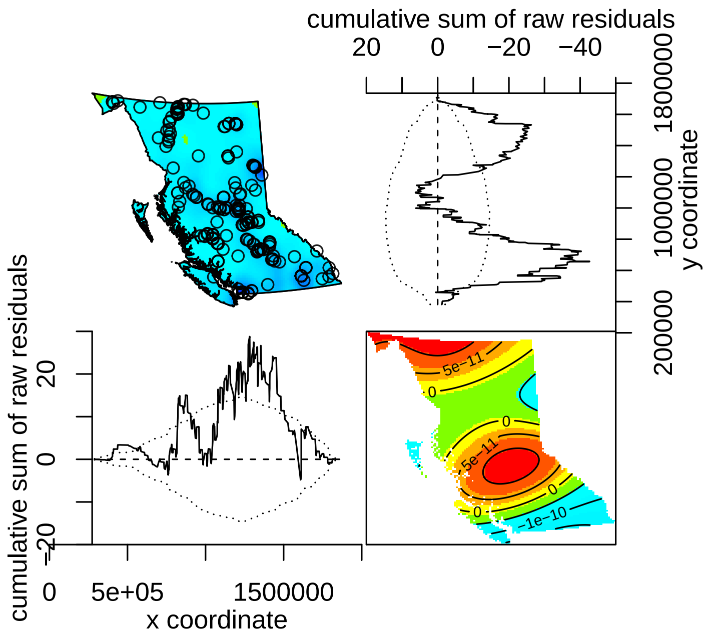




Elevational effect at mean HFI
λ(Elevation)
0.0e+00
5.0e−11
1.0e−10
1.5e−10
0
1000
2000
3000
Elevation
Effect of HFI at mean elevation
λ(HFI)
0e+00
2e−09
4e−09
6e−09
8e−09
0.0
0.2
0.4
0.6
0.8
HFI

# Estimated Red Fox intensity





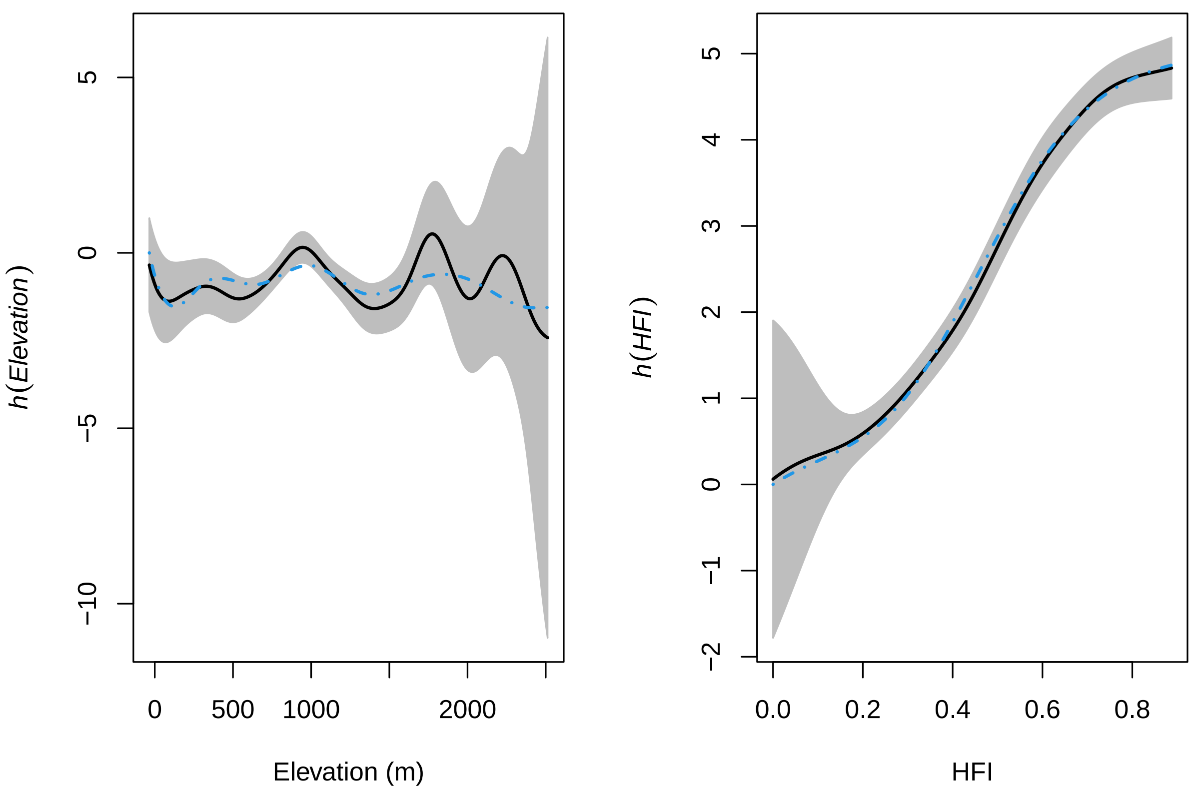

```
## Model diagnostics (raw residuals)
## Diagnostics available:
##  four-panel plot
##  mark plot
##  smoothed residual field
##  x cumulative residuals
##  y cumulative residuals
##  sum of all residuals
## sum of raw residuals in entire window = -5.478e-06
## area of entire window = 9.483e+11
## quadrature area = 9.39e+11
## range of smoothed field =  [-1.561e-10, 1.393e-10]

## Nonstationary Poisson process
## Fitted to point pattern dataset 'parks_ppp'
##
## Log intensity:  ~bs(Elevation, 12) + bs(HFI, 5)
##
## Fitted trend coefficients:
##         (Intercept)  bs(Elevation, 12)1  bs(Elevation, 12)2  bs(Elevation, 12)3
##        -23.06011643         -2.50238970         -0.03994102         -1.39988574
##  bs(Elevation, 12)4  bs(Elevation, 12)5  bs(Elevation, 12)6  bs(Elevation, 12)7
##          0.22518356         -1.66175832         -0.54356329         -0.57428415
```

```
##  bs(Elevation, 12)8  bs(Elevation, 12)9 bs(Elevation, 12)10 bs(Elevation, 12)11
##         -1.89370408         -1.12420362         -4.85151633          6.54909789
## bs(Elevation, 12)12          bs(HFI, 5)1         bs(HFI, 5)2         bs(HFI, 5)3
##        -29.55917527          0.33782061          0.39850473          4.19857566
##          bs(HFI, 5)4          bs(HFI, 5)5
##          4.73799410          4.86649751
##
## For standard errors, type coef(summary(x))
## Problem:
##  Values of the covariate 'HFI' were NA or undefined at 0.56% (12 out of 2137)
## of the quadrature points
```

```
## [1] -4.563551
```

```
## [1] 10408.13
```




Higher powers for elevation - error with convergence

```
## [1] 10401.62
```

```
## [1] 10403.57
```

```
## [1] 10408.13
```

```
## Analysis of Deviance Table
##
## Model 1: ~HFI + I(HFI^2) + Elevation + I(Elevation^2)     Poisson
## Model 2: ~Forest + I(Forest^2) + HFI + I(HFI^2) + Elevation + I(Elevation^2)      Poisson
##   Npar Df Deviance Pr(>Chi)
## 1    5
## 2    7  2   5.9418  0.05126 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### model: Tablulate observations -4. table to summarize models - linear+quad, AIC

### model: GAM - Table if necessary

## Discussion: Provide a brief summary of your findings. Length: ca. 1 page.

## References: Include references to all necessary literature.


1. Data: GBIF.org (09 April 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.p6tsaa
2. Research topics: https://cwf-fcf.org/en/resources/encyclopedias/fauna/mammals/red-fox.html