Regression methods for constructing species distribution models for eagle-use in the continental United States

Justine Fretz[1], Annaliese Chen[2], Eugene Monforte[3], and Dr. Leslie New[*]

[1] *Lafayette College*
[2] *Swarthmore College*
[3] *Cal Poly Pomona College*
[*] *Ursinus College*

January 12, 2024

1 Abstract

Species distribution models (SDMs) are a statistical approach that uses species abundance and environmental data to predict species' distribution across a landscape or oceanographic area. An SDM can take the form of a generalized linear model, generalized additive model, or kernel density estimation. Our research focuses on using existing SDM frameworks to model and analyze eagle distribution across the continental United States. We will use an SDM to better predict eagle-use at wind facilities in the United States. Our SDM will allow for the extrapolation of existing data to predict the distribution of golden and bald eagles at any proposed wind facility site prior to its construction. These predictions can be used in partnership with the United States Fish and Wildlife Service (Service) to inform permit decisions for wind farms to legally take eagles. Eagle take permits are required for wind farms to operate due to fatal collisions of bald and golden eagles with wind

turbines. Our research seeks to understand the relationship between eagles and wind farm locations while minimizing the time and funding resources associated with the data collection required.

2 Introduction

With climate change becoming a more prominent topic in today's society, there has been greater push for the development and use of renewable energy such as wind, wave, and solar power [1]. The development of these technologies is driven by the desire to produce energy without the effects of burning fossil fuels or polluting the air [1]. However, no technology is without cost, with the conflicts between wind energy development and wildlife conservation being particularly notable. Of great concern are golden eagles (*aquila chrysaetos*) and bald eagles (*haliaeetus leucocephalus*) which are known to collide with wind turbines causing death or unsustainable injury (CITE). Under The Bald and Golden Eagle Protection Act (BGEPA), these birds have federal protections placed upon them, which prohibit any person without a permit to harass, harm, possess, or kill them (Eagle Act; 16 U.S.C 668-668d). To try and mitigate the effect of wind facilities on eagles in the US, the US Fish and Wildlife Service (the Service) has instituted what is known as the Eagle Take Permit.

The Eagle Take Permit uses a Bayesian Collision Risk Model [2] to model the relationship between eagles' use of an area prior to construction and the ensuing fatalities of birds per year, taking into account a wind facility's hazardous footprint and the probability of a collision. The manual collection of site-specific data related to eagle-use of the landscape on which a wind facility is located is often perceived as time consuming and costly. This has limited the availability of the data required to perform statistical analyses. As a result, there is increased interest in the utilization of predictive modeling. The primary cost associated with manual data collection is time, as some projects can take multiple years to obtain sufficient information regarding eagle space use. Prior to construction, individual observers have to routinely watch the skies and monitor the time eagles spent flying within a defined spatial and temporal extent. This collection of eagle-use data is used to predict bird fatalities prior to construction. Wind facilities also have to conduct post construction carcass monitoring, which requires observers to habitually walk transects to report the number of eagle carcasses located near wind turbines. This enables researchers to be able to estimate the actual number of eagles to have collided with wind turbines.

The use of predictive modeling can help eliminate some of the perceived costs of manual data collection and potentially allow for more informative applications of the collision risk model when data are not available. Predictive modeling is a process that is used to predict the outcome of an event by taking into account preexisting data and analyzing the patterns. It has been used in the medical field where known patterns in response to a treatment of patients with inflammatory bowel disease matched a model that would have predicted similar results [3]. In an ecological setting, predictive modeling has been used to mediate animal-vehicle collisions [4]. It has also been used for predicting the amount of feed housed cattle consume based on general characteristics about the cattle [5]. One type of predictive modeling whose use has been growing are Species Distribution Models (SDMs), which use environmental covariates to create a model of the species occurrence or density across a defined spatial extent. Once a model is developed, it can be used estimate or make predictions of a species across a larger area or under different environmental conditions. SDMs can aid in conservation efforts and wind facility permitting decision-making processes by providing insights into species' spatial responses to environmental changes.

For eagles, SDMs have been used to estimate the probability of eagle nest site occurrence based on topological, climatic, and land use data. These studies have shown that altitude and slope were significant predictors contributing to eagle nest occurrence in an area [6]. Another significant factor contributing to the occurrence of eagle nests is the diverse regions golden and bald eagles are known to occupy across the continental United States, including North American Deserts, Eastern Temperate Forests, and Great Plains. As a result, when considering predictive models for eagle-use, the climate, topology,and regions will be analyzed for potential relationships that could provide more insight into the preferences and habits of eagles.

Using data from wind facilities from across the contiguous United States, we will be using generalized additive mixed models (GAMMs) to build species distribution models for bald and golden eagles to determine what environmental factors influence their use of the landscape surrounding wind facilities. In this paper, we will be showcasing the results and methodology from our analysis of the relationship between eagle-use of the landscape at wind facilities and data on the climate, eagle relative abundance, and topology of the continental United States. Using SDMs, we will discuss how certain environmental factors effect the use of an area actively being inhabited by eagles. The model will be used to create spatial maps which show predicted eagle-use. This will aid informing the Service's Eagle Take Permits in situations where data are not available, and thus help in determining whether or not a wind facility should be constructed at a proposed site.

3 Methods

3.1 Data

The data for the analysis was provided by the Service and included 198 observations from 119 proposed and constructed wind facilities from 28 states. Of those 198 observations, 101 had information on bald eagles, and 97 had information on golden eagles. Given the uniqueness of the environmental conditions in Alaska compared to the contiguous U.S., the three wind facilities in the data set from that state were excluded from the analysis, resulting in 195 observations from 117 wind facilities, 99 had information on bald eagles, and 96 had information on golden eagles. A portion of our data set consists of predicted eagle minutes using XXX model as developed in *New et al., 2021*[2]. Another form of eagle data that was used as an explanatory variable in our analyses was the species' relative abundance. This was defined based on the eBird relative abundance data for each species. Golden and bald eagle data was modeled separately, given the different life history characteristics and ecology of the two species.

The environmental variables chosen to potentially inform the models reflect many of the choices made in Dunk et al. **Table ??** lists the the chosen covariates. We made the decision to include the level 1 eco-region over the level two eco-region due to the size of our data set. The level 2 eco-region is over-specific about every location, and due to the nature of our data set, we would have level 2 eco-region variable locations with very few observations. While latitude and longitude where available for each wind facility, eagles are responding to the landscape at a wider scale. Therefore a resolution and buffer radius had to be selected. This also helps to ensure consistent data extraction processes. A 10 kilometer buffer was selected to account for the variation across an area that that would encompass the footprint of the majority of operating wind facilities.

The 10km buffer was created around the points latitude and longitude

Covariates Chosen from Dunk et al.			
Variable Name	Units	Variable Description	Used in Final Model
adi	degree-days (#)/millimeter (mm)	Annual Dryness Index	X
mmi	degree-days (#)/millimeter (mm)	Annual Moisture Index	X
baea_rel_n		Bald Eagle Relative Abundance	
dd5	degree-days (#)	Degree Days > 5°C	X
goea_rel_n		Golden Eagle Relative Abundance	
led	meters (m)	Local Elevation Difference	X
map	millimeters (mm)	Mean Annual Precipitation	X
na_l1name		Ecoregion Level I	
na_l2name		Ecoregion Level II	
oro	meters (m)/second (s)	Orographic Uplift	X
region		Region of the US in which the wind facility is found	
ther	meters (m)/second (s)	Thermographic Uplift	X
tri	meters (m)	Terrain Ruggedness Index	X
tri2	meters (m)	Topographical Ruggedness Index	X
twi	meters (m)	Topographical Wetness Index	X
twnd	meters/second (m/s)	Total Wind Speed	

Table 1: X denotes variable values for sites calculated using rasters from Dunk et al. 3km resolution with an average of values within a 10km buffer of the coordinates listed for the facility

in the exposure data set. The values that lie within the buffer region for each point were averaged and the value was used as an explanatory variable in the SDM. In order to ensure the process was accurately extracting the data, a check was set in place to ensure the proper size buffer was created as well as to ensure the values were reasonable given the range of possible values for each covariate. This process required particular attention to the coordinate reference system (CRS) to ensure the data was being extracted from the proper location.

The land cover data set was collected from the Commission for Environmental Cooperation [7]. The data set was originally defined by 19 land cover categories. However, to reduce the amount of variables in the model, land cover variables were simplified down to type of land cover regardless of temperature zone. For example, the land covers defined as "Temperate or sub-polar needle-leaf forest", "Tropical or sub-tropical broad-leaf evergreen forest", "Tropical or sub-tropical broad-leaf deciduous forest", "Temperate or sub-polar broad-leaf deciduous forest", and "Mixed Forest" were reduced simply to "forest". The variables "water" and "ice" were also merged to one land cover type, "water", as only one wind facility had a minuscule proportion of water. The result of this reduction was 8 variables that were used to inform the SDM. Given that the buffer regions around each wind facility would contain multiple land cover types, the land cover data was transformed into a numerical proportion. The resulting land cover variables represented the proportion of each land cover type at a 3km resolution across the United States. Then, the buffer was extracted as described above. Given that the maps were at a 3km resolution, while the buffer region was a 10km radius, the values within the region were averaged to create the covariate used in the SDM.

In order to create a data set to be used for model predictions, we had to compile the covariate data from across the contiguous United States, ensuring all files used the same coordinate reference system and had the same resolution and scale. These raster files were then stacked to create a Raster Stack with each layer representing a different covariate used in the model. It is important to note that the layer names of the Raster Stack must match the variable names of the data set used to create the model. The raster stack can then be used along with the model to predict eagle-use across the contiguous United States. Plotting these predictions creates a heat map of the response variable in this case eagle minutes per unit effort based on the specified models.

3.2 Model

We used an SDM to create a spatial mapping of eagle-use in the contiguous U.S., using the data described above. In modeling our data, we chose to use a generalized additive mixed model (GAMM) because it loosens the assumption of linearity. By doing this, the splines that are part of GAMM allow for more complex models to be fit, giving greater flexibility to the potential relationship between the response and explanatory variables. The ecological region of North America in which the wind facility was located was used as a random effect. We also incorporated a Tweedie distribution to account for the point mass around 0 that could be seen in a histogram of our response variable, eagle-use, defined as eagle minutes per unit effort.

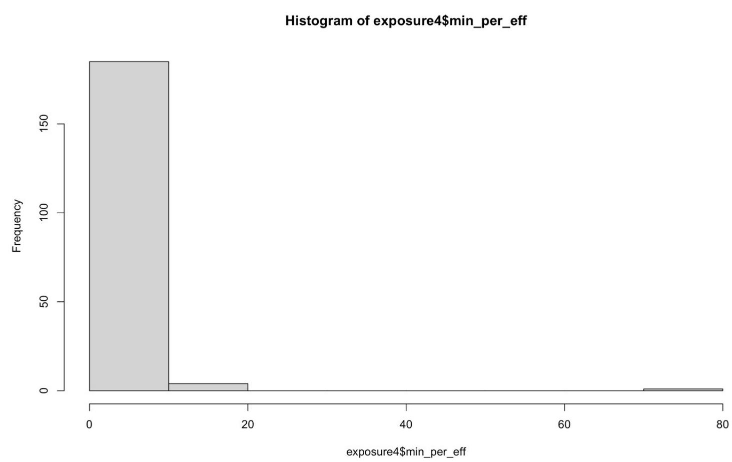

Figure 1: Histogram of Tweedie distribution

$$y_i = \alpha + \sum_j f_j(x_{ji}) + \epsilon_i, \epsilon_i \sim TW_p(\mu, \sigma^2)$$

The y is the response variable. α represents the intercept. The function f represents the spline of each explanatory variables x. ϵ is the error. $\epsilon \sim TW_p$ represents that we are using a tweedie link function and p is the parameter that controls the shape of the tweedie. μ represents the mean and σ represents the variance.

3.3 Analysis

Exploratory data analysis was used to investigate potential relationships between variables. Explanatory variables that had a correlation higher than ± 0.80 were investigated more closely, and one of the variables removed from the potential set to avoid issues with collinearity.

Exploratory data analysis for bald eagles showed that annual dryness index and annual moisture index were highly correlated (.955). For bald eagles, mean annual precipitation and annual moisture index were correlated (.885). Topographical ruggedness index and terrain ruggedness index were also correlated. Topographical wetness index was correlated to topographical ruggedness index (.873) and terrain ruggedness index(.870).

Exploratory data analysis for golden eagles showed that annual dryness index and annual moisture index were correlated (.975). Thermal uplift and mean annual precipitation were correlated (.883). Topographical ruggedness index and terrain ruggedness index were correlated (.998). Topographical wetness index was correlated to topographical ruggedness index (.899) and terrain ruggedness index(.895). Shrubland and thermal uplift were correlated (.813).

Covariates that were correlated were not modeled together.

After the exploratory data analysis, the largest reasonable model for golden eagles included the variables XXX, while the largest reasonable model for bald eagles included YYY. These models were checked for concurvity before proceeding with any other model fitting. When two variables are concurve, one variable may be a smooth curve of another, and we should attempt to avoid modeling them together. After checking concurvity, we made decisions as to if one variable should be modeled with another.

The analysis was carried out in the statistical computing language R [8]. Creation of the data set was done using packages raster [9], sp [10], rgdal [11], dplyr [12], ggplot2 [13], readr [14], sf [15], terra [16], tidyverse [17], viridis [18], spData [19], colorRamps [20], rgeos [21], tmap [22], janitor [23], stringr [24], and ggbeeswarm [25]. (CITE all packages), and exploratory data analysis was carried out using *GGally* [26]. All models were fit using the *mgcv* package [27] in the statistical computing language R.

4 Results

We began our data analysis with checking for correlation between variables. The topographic wetness index (twi) was highly correlated with terrain ruggedness index (tri) and topographic ruggedness index (tri2) (r=-0.88 and -0.876, respectively). The variables tri and tri2 were also closely correlated (r=0.998). We chose to keep tri because we referenced the Dunk et. al and they used the tri over tri2. The other notable correlation we found were annual moisture index (ami) and annual dryness index (adi) (r=0.967). We retained XXX for use as a covariate because YYY.

When checking for concurvity, an interesting observation was that explanatory variables that were concurve for one species, were not necessarily concurve for the other.

In our initial model for bald eagles, we identified concurvity among [list the variables here, and refer readers to the table you are creating]. Of these variables, we chose to keep XX because YY. [Continue in this vein]

After our results with concurvity for bald eagles, we decided to not use wetland since it was concurve with bald eagle relative abundance. We expect relative abundance to be significant in our model, thus making the decision to choose the relative abundance rather than wetland. We also found that degree days over 5 Celsius was concurve to urban. Water was also concurve to degree days over 5 Celsius. Thus, we decided to not use degree days over 5 celsius. Since we predict water to have a big effect with bald eagles due to them being open-water foragers. Mean annual precipitation and ecoregion level 1 were concurve so we decided to not use mean annual precipitation since we want ecoregion to be accounted for in our largest model. Ecoregion should be accounted for since we expect golden and bald eagles to act differently in different areas of the United States. Urban and water ice are concurve so we would have to have a model with only one of them. Annual dryness index and annual moisture index were not used because they were concurve with ecoregion level 1. Local elevation difference was concurve with topographic wetness index and total ruggedness index. Our largest model would only consist of one of these. The largest model we decided to use consisted of the explanatory variables bald eagle relative abundance, ecological region, water, total wind speed, orographic uplift, topographic wetness index, urban, cropland, and mean annual precipitation. See Table 4.

Using AIC for our model selection process, we found that, for bald eagles, the best model in the set was the one that included bald eagle relative

Table 2: Covariates of Largest Bald Eagle Model

	s(baea_rel_n)	s(relevel)	s(water)	s(twnd)	s(oro)	s(twi)	s(urban)	s(cropland)	s(map)
s(baea_rel_n)									
s(relevel)									0.8
s(water)							1		
s(twnd)									
s(oro)									
s(twi)									
s(urban)									
s(cropland)									
s(map)									

Table denoting all the variables used in the Golden model for Bald Eagles and their concurvity.

Table 3: 10 Best Bald Eagle Delta-AIC

Model Name	Δ AIC
gam(min_per_eff ~ s(baea_rel_n)+s(relevel, bs="re")+s(water_ice)+s(twnd)+s(oro)+s(twi), family=tw, method = "ML")	0
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice) +s(twnd)+s(oro)+s(twi)+s(map), family=tw, method = "ML")	1.13
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice) +s(twnd), family=tw, method = "ML")	1.26
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice) +s(twnd) +s(oro)+s(twi)+ s(cropland), family=tw, method = "ML")	1.89
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice) +s(twnd) +s(oro), family=tw, method = "ML")	1.96
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice), family=tw, method = "ML")	1.99
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice)s(oro),family=tw, method = "ML")	2.71
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice) +s(twnd)+s(twi), family=tw, method = "ML")	2.89
gam(min_per_eff ~ s(baea_rel_n) + s(relevel, bs="re") + s(water_ice)+s(twi), family=tw, method = "ML")	3.4
gam(min_per_eff ~ s(baea_rel_n)+s(relevel, bs="re")+s(water_ice)+s(twnd)+s(oro)+s(map), family=tw, method = "ML")	4.17

Table denoting the 10 best models for Bald Eagle model and comparing their AIC to the lowest one

abundance, ecological region, water ice, total wind speed, orographic uplift, and topographic wetness index as explanatory variables.

In our initial models for bald eagles, we identified concurvity among [list the variables here, and refer readers to the table you are creating]. Of these variables, we chose to keep relative abundance over mean annual precipitation because we expect as relative abundance increases, eagle-use will also increase. We chose to keep ecoregion level 1 over thermal uplift and shrubland since we want to take in account ecoregion since we expect different regions. We chose to use orographic uplift over local elevation difference since eagles like places with high orographic uplift, so we expect to see more eagle-use in those areas. We chose to not use annual moisture index and annual dryness index since these covariates were concurve to multiple other covariates including ecoregion. We made 4 models that considered variations of either using degree days over 5 Celsius or total wind speed and topographic ruggedness index or topographic wetness. We made 4 models that considered variations of either using degree days over 5 Celsius or total wind speed and topographic ruggedness index or topographic wetness. The largest reasonable model we fit to golden eagles included the covariates: golden eagle relative abundance, ecological region, orographic uplift, topographic ruggedness index, degree days over 5 Celsius, grassland, cropland, forest, and urban.

After our results with concurvity for golden eagles, we found that golden eagle abundance and mean annual precipitation to be concurve. We decided to not use mean annual precipitation since we expect golden eagle relative abundance to play a big role in our models. Annual moisture index and annual dryness index were both concurve with multiple variables, so we decided to not include those variables in our model. Topographic wetness index and topographic ruggedness index were concurve with each other, so only one would be in our largest model. Thermal uplift wasn't used due to it being concurve with ecoregion level 1 and we want to account for ecoregion in our largest model. We chose orographic uplift over local elevation difference since local elevation difference was really close to being concurve to topographic wetness index and topographic ruggedness index. Shrub-land was concurve to grassland, ecoregion level 1, and cropland, so we decided to not use shrubland. We made 4 models that considered variations of either using degree days over 5 Celsius or total wind speed and topographic ruggedness index or topographic wetness. The largest reasonable model we fit to golden eagles included the explanatory variables golden eagle relative abundance, ecological region, orographic uplift, topographic ruggedness index, degree days over 5 Celsius, grassland, cropland, forest, and urban.

Using AIC for our model selection process, we found that, for golden eagles, the best model in the set was the one that included golden eagle relative abundance, ecological region, degree days over 5 Celsius and grassland, cropland and forest landcover percentage.

The final model for bald eagles had an adjusted r^2 of .931 and a deviance explained of 80.4 %. The shape of the smooth of bald eagle relative abundance was linear and increasing. Water has a smooth that increases until about where water is 0.2, then starts to decrease. Total wind speed has a smooth that increases, then decreases, then increases again. The orographic uplift has a smooth that increases to about -.01 before decreasing. Topographical wetness index has a smooth that is linear and decreasing.

Table 4: Covariates of Largest Golden Eagle Model

	s(goea_rel_n)	s(oro)	s(twi)	s(dd5)	s(relevel)	s(shrubland)	s(grassland)	s(urban)	s(cropland)	s(forest)
s(goea_rel_n)										
s(oro)									0.8	
s(twi)							1			
s(dd5)										
s(relevel)						0.8				
s(shrubland)							0.84		0.81	
s(grassland)										
s(urban)										
s(cropland)										
s(forest)										

Table denoting all the variables used in the largest model for Golden Eagles and their concurvity.

Table 5: 10 Best Golden Eagle Delta-AIC

Model Name	Δ A
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(dd5)+s(grassland)+ s(cropland), family= tw, method= "ML")	0
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro) + s(dd5)+s(grassland)+ s(forest), family= tw, method= "ML")	.1
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re") + s(dd5)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	.3
gam(min _per _eff ∼ s(goea _rel _n) + s(dd5)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	.5
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(dd5)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	1.7
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(tri)+s(dd5)+s(grassland)+s(forest), family= tw, method= "ML")	1.9
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	2.0
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(tri)+s(dd5)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	3.7
gam(min _per _eff ∼ s(goea _rel _n)+ s(relevel, bs="re")+ s(oro)+s(dd5)+s(cropland)+s(forest), family= tw, method= "ML")	5.9
gam(min _per _eff ∼ s(goea _rel _n)+s(grassland)+s(cropland)+s(forest), family= tw, method= "ML")	6.9

Table denoting the 10 best models for Golden Eagle model and comparing their AIC to the lowest one

The final model for golden eagles had an adjusted r^2 of .651 and deviance explained of 73.7 %. The shape of the smooth of golden eagle relative abundance shows a increasing logarithmic looking curve. The smooths of the degree days over 5 Celsius, cropland, and forest all showed a decreasing line. The smooth of grassland looked like a sine curve where there where the smooth first increases, then decreases, then increases again, before decreasing again.

We checked assumptions for our models using the gam.check function in the mgcv library and both models met all assumptions.

Using the final model to use for each species, we began to predict across the contiguous United States using the model. Once the predictions were made, they were plotted on a map of the country. Additionally, maps were created for the standard error to evaluate the accuracy of our models. See Fig 2, Fig 3, Fig 4, and Fig 5.

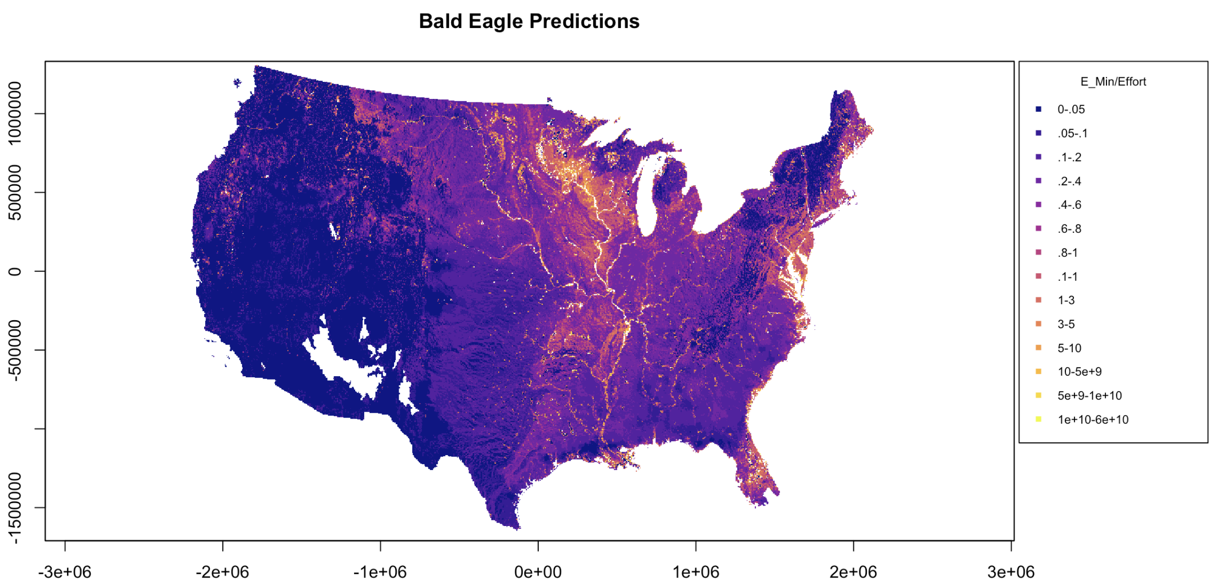

Figure 2: Bald Eagle Predictions

Looking at the maps, it is evident across both species that a majority of predictions are relatively low with a few areas having more extreme values

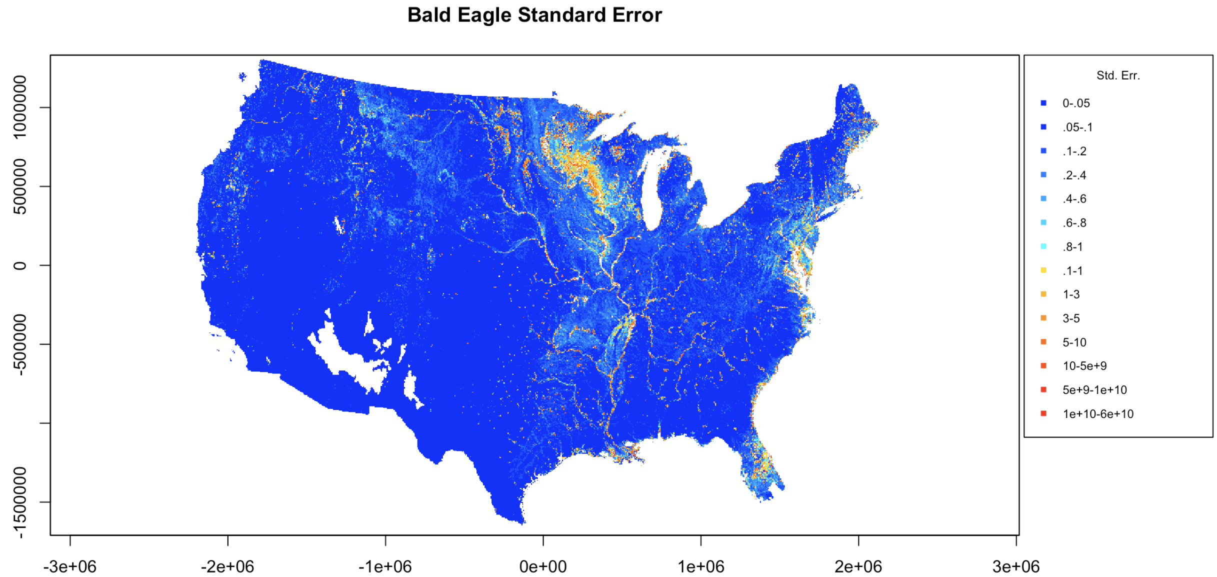

Figure 3: Standard Error for Bald Eagle Predictions

where the values of the covariates tend to be more extreme as well. This is not of particular concern as wind facilities are often not constructed in areas with extreme topographical or climatic variables. However, the lack of differentiation in the predictions across much of the great plains and south-eastern United States compared to that of mid-west is quite evident. This would imply that our model does not give very accurate predictions over areas where data is lacking as there is little to no wind facility data to help inform the model for these areas.

With respect to the Golden Eagle predictions, their presence predominantly in the western half of the United States according to the map, follows with ecological knowledge of Golden Eagle presence in that region. This may also explain some of the lack of differentiation in predictions across the eastern half of the United States. The Bald Eagle predictions were quite different comparatively. As a result of water being included in the model, we observed extremely high predictions along water ways. This does however follow the knowledge of Bald Eagle presence predominantly around larger water sources. In the following section, the potential reasons behind these

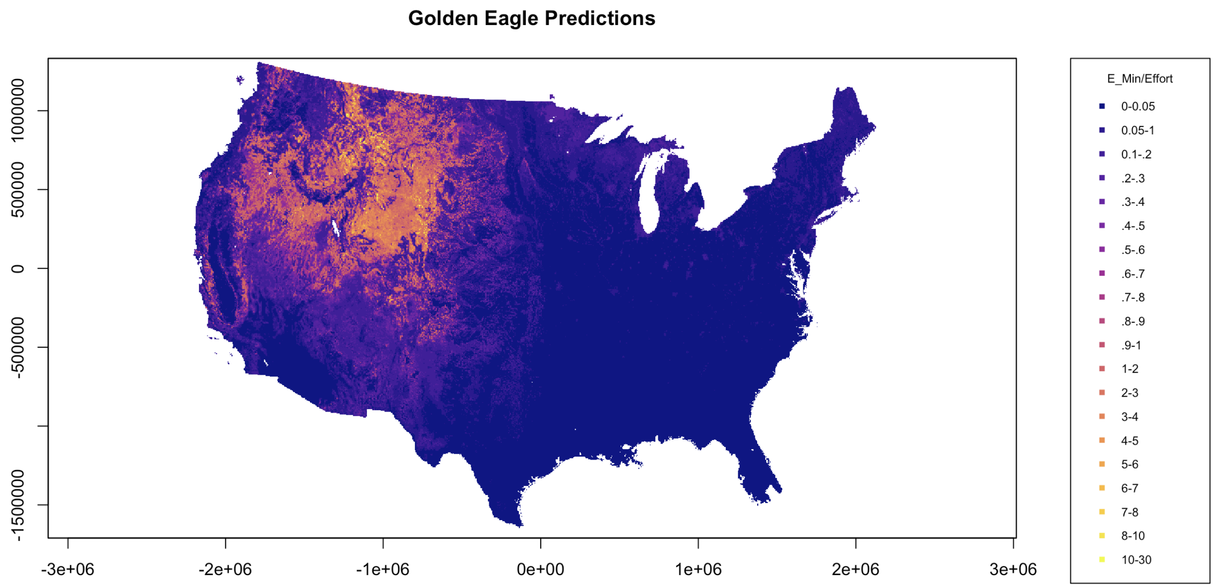

Figure 4: Golden Eagle Predictions

results and their connection to both applications and areas of further study will be discussed.

5 Discussion

The goal of our analysis was the prediction of eagle-use in places where actual data observations were not collected. The goal of this project is to inform the Service and its policies regarding Eagle Take Permits and future eagle conservation research.

While we considered a large set of potential covariates, our data set was limited to only 195 observations in total, with under 100 for the individual eagle species. This poses an issue, as too many explanatory variables with a small data set leads to over-fitting. To account for this, smaller groups of covariates were created through a careful selection process. Models were fit to the smaller subsets of covariates, but the limitation imposed by the small sample size means that our models may not accurately account for all

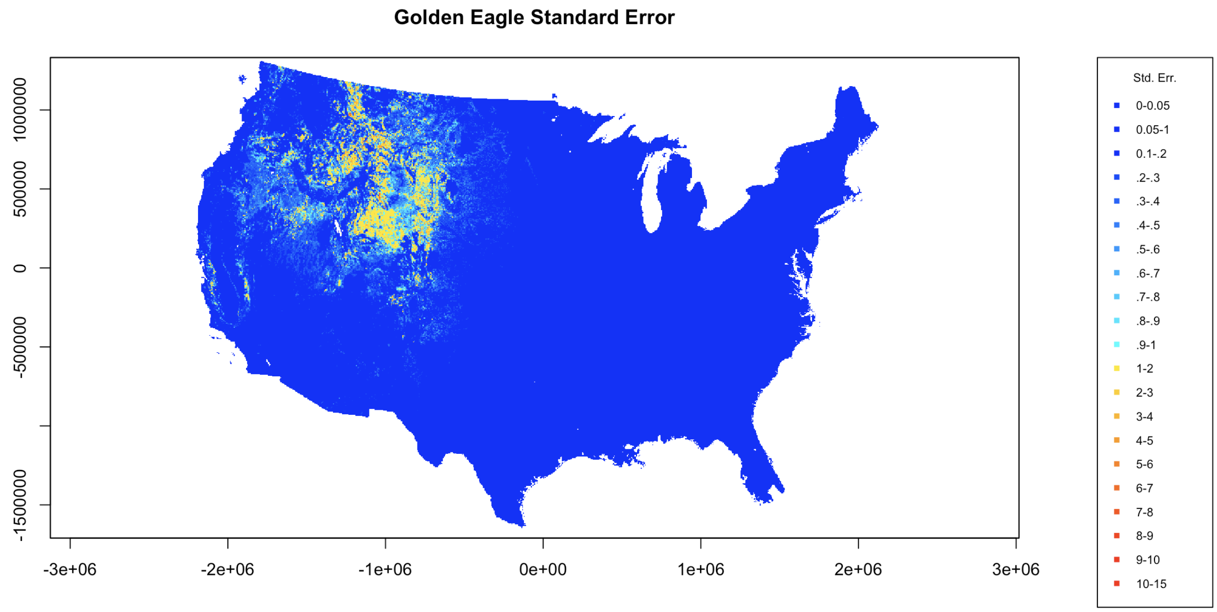

Figure 5: Standard Error for Golden Eagle Predictions

factors influencing eagle-use. A remedy to consider would be to obtain data from more wind facilities across the United States, which would allow for an increase in the number of variables we could incorporate into our model. The small sample size can influence the accuracy of our model, which can translate into issues with respect to extrapolation of our predictions as we project the model across the United States. Since certain regions of the country were underrepresented in our sample, areas with little to no data boasted larger amounts of uncertainty in the model predictions. To further explore this, it would be beneficial to consider the creation of isolated models based on the region of the country where some variables, such as climate and topography, are less varied, which might, provided a sufficient sample size was available, help reduce the uncertainty in the predictions.

Another avenue for further research would be to consider the effects of varying the buffer region created around each wind facility observation. The size of the buffer regions were determined relative to size of wind facilities in the U.S., as this would encapsulate the area over which the data were collected at each site. However, this may not be the scale at which eagles are responding to characteristics of the landscape or climate. Given that eagle-use is averaged across space and time at each wind facility, it is unlikely that a buffer smaller than 10 km would be informative, but the effect of choosing a larger buffer should definitely be explored.

It would be worthwhile to consider the fact the data has some level of bias due to being collected from locations across the United States where wind facilities have been proposed to be built. Therefore, locations featuring high wind speed and open areas which are features that would make for a prime wind facility location, are more present in our model which would suggest some bias as our model is extrapolated into those regions.

Ideally, our models would have incorporated information on prey availability, as this is considered a major driver of eagle distribution [28], but prey availability data are not available at a national scale. However, our use of ecological regions in North America as a random effect should capture some aspects of possible prey availability on a broad scale. As a suggestion, when looking at a possible wind facility location, it would be advantageous to consider surveying the area for prey, as it is known that an abundance of prey with likely lead to an increase in eagle-use [28].

The spline of bald eagle relative abundance shows that as relative abundance increase, we see eagle-use increased. The spline of water ice shows that eagle-use increases with water ice until about 0.2 and then eagle-use starts to decline after that. The water ice curve is heavily influenced by the small amount of data with this landcover. We see the confidence interval start to increase at around 0.2 water ice. The curve is heavily influenced by points after 0.1 water ice. Total wind speed has a spline that shows eagle-use steadily increasing until about 5.5 total wind speed. We see eagle-use decline when total wind speed is between 5.5 and 7. We then see eagle-use start to increase again as total wind speed goes past 7. The confidence interval is really wide past 7 total wind speed, being influenced heavy by the small amount of data we have after that point. The orographic uplift spline shows eagle-use increase until about -.02 orographic uplift. Then we see eagle-use decline after that point. The confidence interval is very wide at points before the peak at -.02, again due to small amounts of data. The terrain wetness index shows a decline in eagle-use as the terrain wetness index increases.

The spline of golden eagle relative abundance shows that as golden eagle relative abundance increases, we see an increase of eagle-use. The majority of our data is from 0 to 3.5 golden eagle relative abundance. After 3.5, we see only two points that influence the curve. The splines of degree days over

5 Celsius, cropland and forest, we see a decreasing linear relationships with eagle-use. The grassland spline increases, decreases, increases then decreases again. This looks like the result of having small amount of data points and the data points being spread out.

CONCLUSION PARAGRAPH

6 Acknowledgements

Huge shout out to Dr. Leslie New for her mentorship throughout this program. Thank you to Todd Lickfett as well as Hillary White, Emily Bjerre, and Paige Howell at the US Fish and Wildlife Service for their collaboration and support. This research was completed as part of the Ursinus College REU Program courtesy of NSF Grant #1851948.

7 References

References

[1] Sanya Carley. "Distributed generation: An empirical analysis of primary motivators". In: *Energy Policy* 37.5 (2009), pp. 1648–1659. ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2009.01.003.

[2] Leslie New et al. "A Collision Risk Model to Predict Avian Fatalities at Wind Facilities: An Example Using Golden Eagles, Aquila chrysaetos". In: *PLOS ONE* 10.7 (July 2, 2015). Publisher: Public Library of Science, e0130978. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0130978. (Visited on 06/08/2023).

[3] Akbar K. Waljee et al. "Algorithms Outperform Metabolite Tests in Predicting Response of Patients With Inflammatory Bowel Disease to Thiopurines". In: *Clinical Gastroenterology and Hepatology* 8.2 (Feb. 1, 2010), pp. 143–150. ISSN: 1542-3565. DOI: 10.1016/j.cgh.2009.09.031. (Visited on 07/13/2023).

[4] Juan E. Malo, Francisco Suárez, and Alberto Díez. "Can we mitigate animal–vehicle accidents using predictive models?" In: *Journal of Applied Ecology* 41.4 (2004), pp. 701–710. ISSN: 1365-2664. DOI: 10.1111/j.0021-8901.2004.00929.x. (Visited on 07/13/2023).

[5] C. Davison et al. "Predicting feed intake using modelling based on feeding behaviour in finishing beef steers". In: *Animal* 15.7 (2021), p. 100231. ISSN: 1751-7311. DOI: https://doi.org/10.1016/j.animal.2021.100231.

[6] Jeffrey R. Dunk et al. "Modeling spatial variation in density of golden eagle nest sites in the western United States". In: *PLOS ONE* 14.9 (Sept. 30, 2019). Ed. by Lyi Mingyang, e0223143. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0223143. (Visited on 06/13/2023).

[7] Mireille Pasos. *North American Land Cover, 2020 (Landsat, 30m)*. Commission for Environmental Cooperation. URL: http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/ (visited on 07/19/2023).

[8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023.

[9] Robert J. Hijmans. *raster: Geographic Data Analysis and Modeling*. R package version 3.6-23. 2023.

[10] Roger S. Bivand, Edzer Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013.

[11] Roger Bivand, Tim Keitt, and Barry Rowlingson. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.6-7. 2023.

[12] Hadley Wickham et al. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.2. 2023.

[13] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4.

[14] Hadley Wickham, Jim Hester, and Jennifer Bryan. *readr: Read Rectangular Text Data*. R package version 2.1.4. 2023.

[15] Edzer Pebesma and Roger Bivand. *Spatial Data Science: With applications in R*. Chapman and Hall/CRC, 2023, p. 352.

[16] Robert J. Hijmans. *terra: Spatial Data Analysis*. R package version 1.7-29. 2023.

[17] Hadley Wickham et al. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.

[18] Garnier et al. *viridis(Lite) - Colorblind-Friendly Color Maps for R*. viridis package version 0.6.3. 2023. DOI: `10.5281/zenodo.4679424`.

[19] Roger Bivand, Jakub Nowosad, and Robin Lovelace. *spData: Datasets for Spatial Analysis*. R package version 2.3.0. 2023.

[20] Tim Keitt. *colorRamps: Builds Color Tables*. R package version 2.3.1. 2022.

[21] Roger Bivand and Colin Rundel. *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.6-3. 2023.

[22] Martijn Tennekes. "tmap: Thematic Maps in R". In: *Journal of Statistical Software* 84.6 (2018), pp. 1–39. DOI: `10.18637/jss.v084.i06`.

[23] Sam Firke. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. R package version 2.2.0. 2023.

[24] Hadley Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.0. 2022.

[25] Erik Clarke, Scott Sherrill-Mix, and Charlotte Dawson. *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. R package version 0.7.2. 2023.

[26] Barret Schloerke et al. *GGally: Extension to 'ggplot2'*. R package version 2.1.2. 2021.

[27] S.N Wood. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC, 2017.

[28] David W. Stephens, Joel S. Brown, and Ronald C. Ydenberg. *Foraging: behavior and ecolog*. University of Chicago Press, 2007.

8 Visuals to put elsewhere

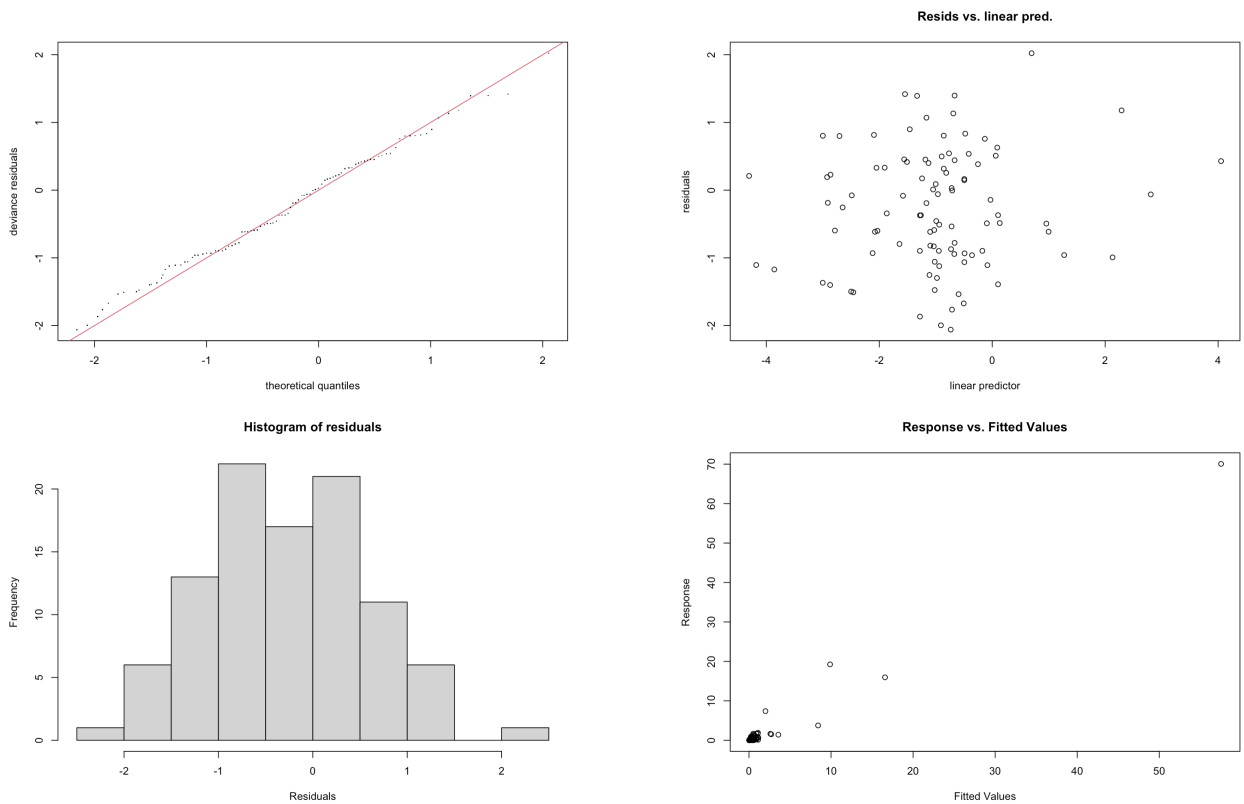

Figure 6: Bald Eagle Model

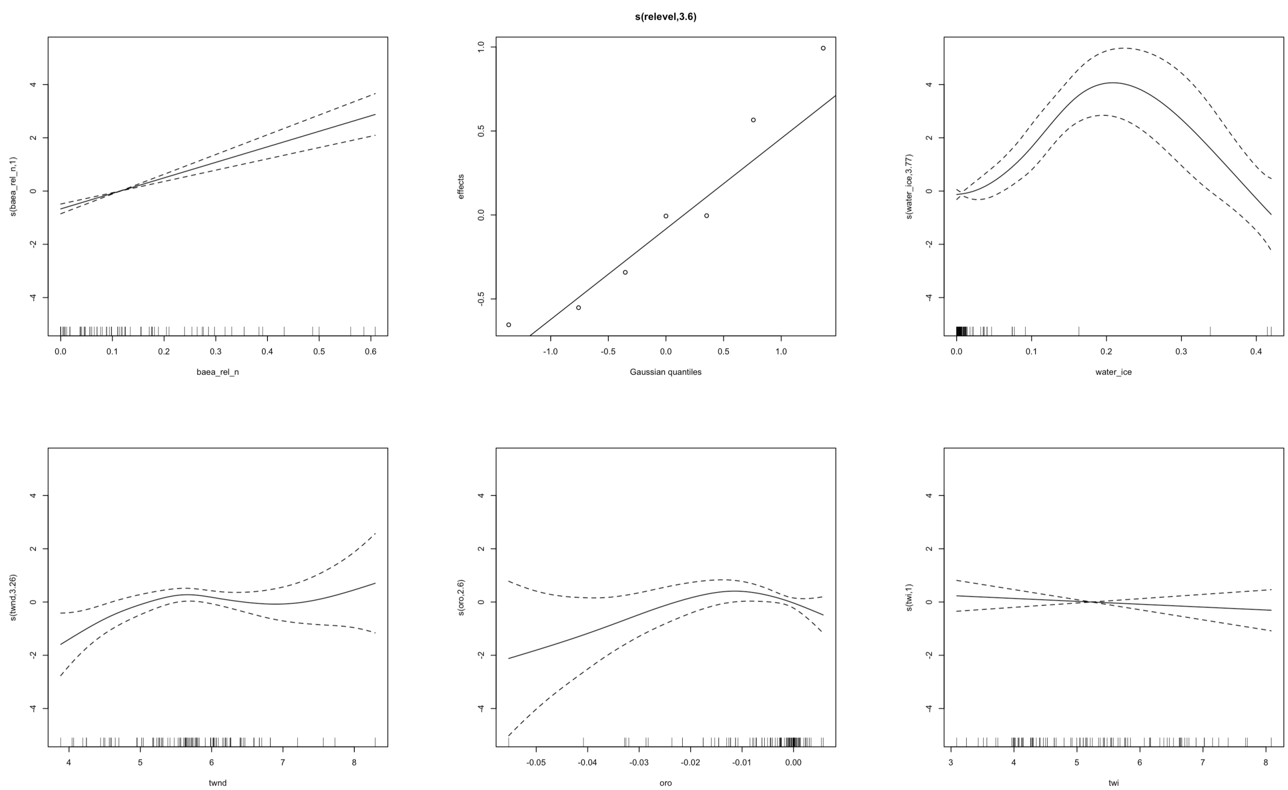

Figure 7: Splines for the Bald Eagle Model

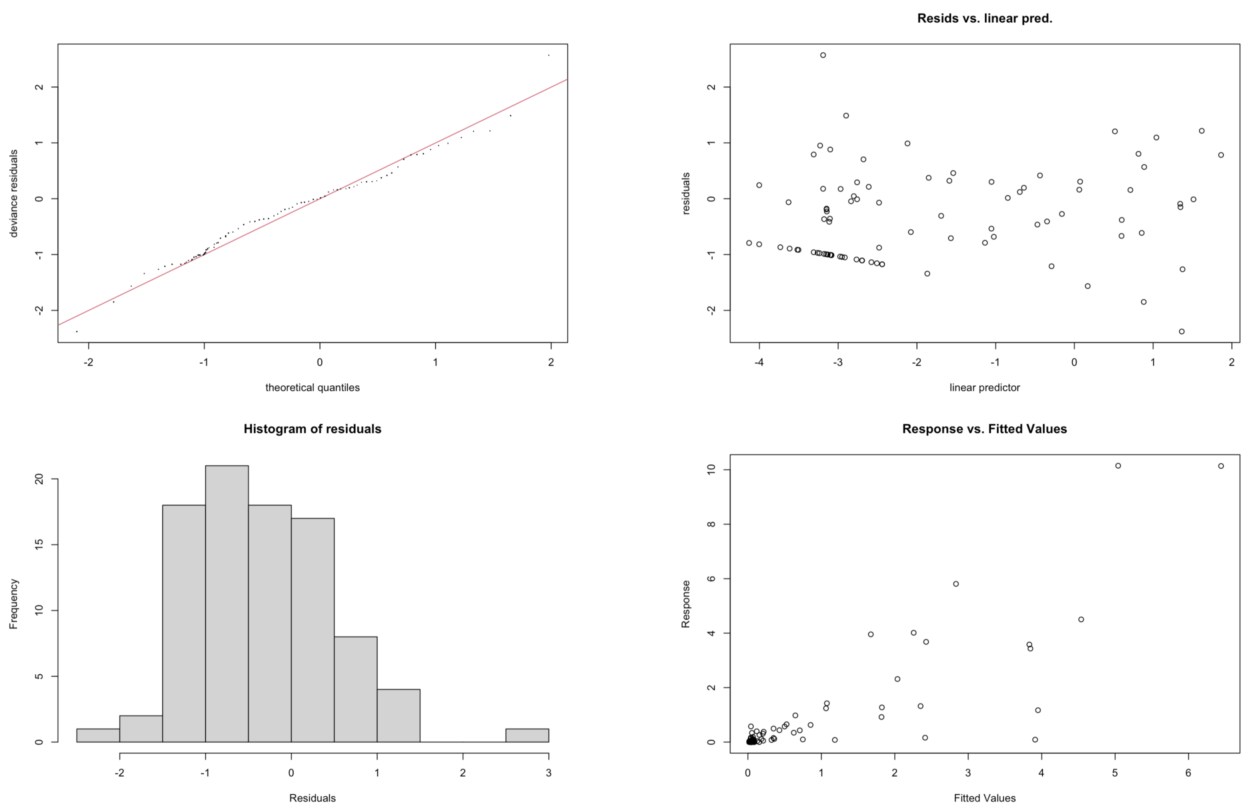

Figure 8: Golden Eagle Model

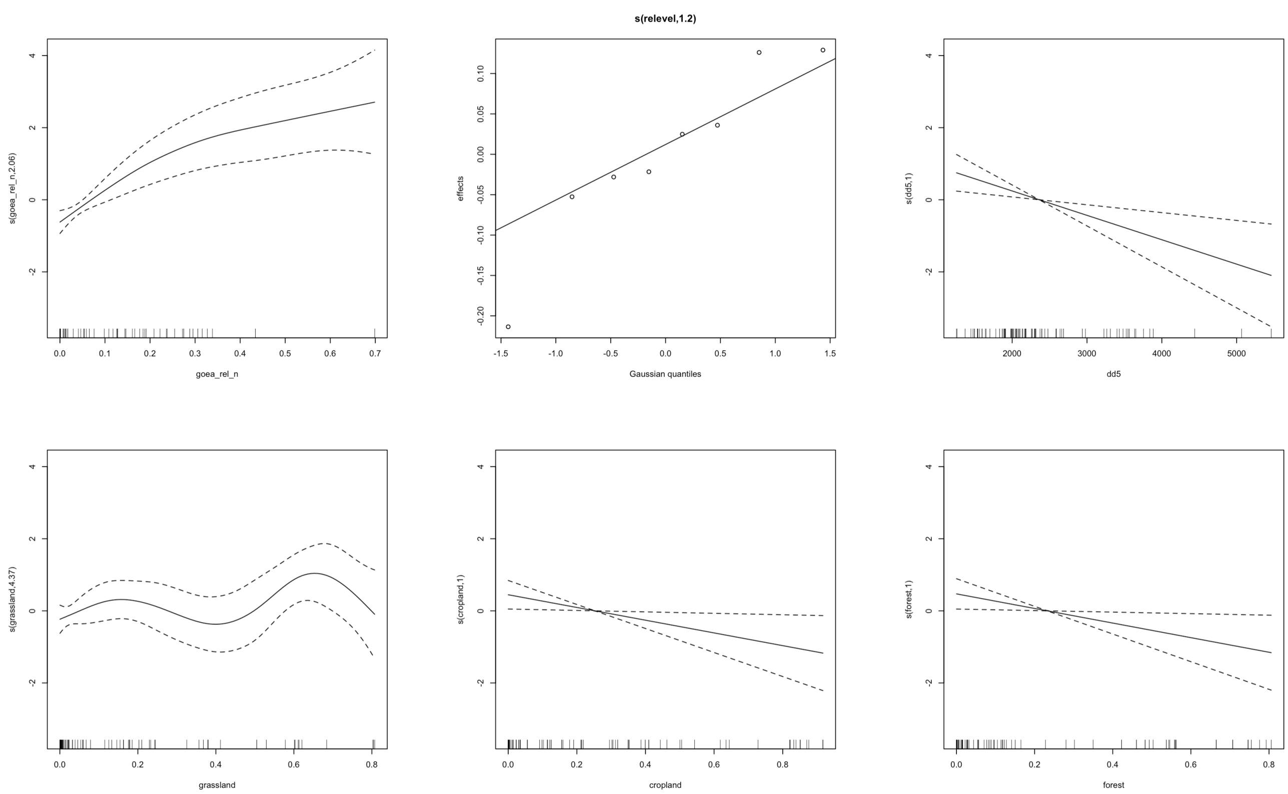

Figure 9: Splines for the Golden Eagle Model