



Selecting priority areas from diversity and individual species abundance **DiversityOccupancy**

Derek Corcoran
University of Missouri

Elisabeth Webb
University of Missouri, USGS Institute for Bird Populations

Dylan Kesler

Abstract

Lately occupancy modeling has been widely used as a tool for ecological research and management planning, most often in the context of single species models. We present the package **DiversityOccupancy** in the R environment. The objective of this package is to simultaneously model factors associated with occupancy and abundance of individual species using a detection history file, and to use predicted abundances to calculate species diversity for each sampling site. The package then models factor(s) associated with among-site species diversity, which can be combined with spatial data to identify areas that contain both high abundance of species of conservation concern and high species diversity.

Keywords: **DiversityOccupancy**, Occupancy Modeling, R.

1. Introduction

1.1. Single-species or multiple-species management

The ecological and management literature has historically valued the idea of managing for diversity. This, however, clashes with the classic way in which conservation takes place. Conservation agencies, governments, scientists and international organizations such as International Union for Conservation of Nature (IUCN), classify species according to a conservation status, and policies are enacted to safeguard the ones that species of conservation concern (Keller and Bollmann 2004; Rodrigues, Pilgrim, Lamoreux, Hoffmann, and Brooks 2006). Managing for single species is often less complicated than simultaneously addressing suites of species, and it is may be simpler to track the status of a species than it is to define management for diversity (Simberloff 1998). Further, tracking the conservation status of a single species is often more straightfroward than tracking multiple species or diversity.

There are several approaches to designing management measures based on single species, such as using umbrella species (Crosby, Elmore, Leslie, and Will 2015; Bichet, Dupuch, Hébert, Le Borgne, and Fortin 2016), Habitat Suitability Indices (HSI) (Reza, Abdullah, Nor, and Ismail 2013; Soniat, Conzelmann, Byrd, Roszell, Bridevaux, Suir, and Colley 2013; Zohmann, Pennerstorfer, and Nopp-Mayr 2013) and Species distribution modeling (SDM) (Peterson 2011; Guisan, Tingley, Baumgartner, Naujokaitis-Lewis, Sutcliffe, Tulloch, Regan, Brotons, McDonald-Madden, Mantyka-Pringle *et al.* 2013) among others. Even when the literature mostly criticizes the use of such approaches due to the ineffectiveness in preventing loss of biodiversity (ROBERGE and Angelstam 2004; Branton and Richardson 2011).

One of the problems of managing simple species, is that is usually very difficult to predict what will happen to other species and because of that it is hard to predict what will happen to diversity (Pulliam 2000), with some examples showing that even management measurements that use some species as umbrella species for conservation of ecosystems leading to undesired community effects (White, Zipkin, Manley, and Schlesinger 2013).

Managing for multiple-species, although desirable, it has shown to be difficult to implement (Möllmann, Lindegren, Blenckner, Bergström, Casini, Diekmann, Flinkman, Müller-Karulis, Neuenfeldt, Schmidt *et al.* 2014; ?), most multimodel species only take into account only the number of species in an area, not accounting for how rare or common some species are, or they do not take into account the presence of some endangered species, since they are only trying to account for higher number of species, and not their identity (Taft, Colwell, Isola, and Safran 2002; Tori, McLeod, McKnight, Moorman, and Reid 2002; Plagányi, Punt, Hillary, Morello, Thébaud, Hutton, Pillans, Thorson, Fulton, Smith *et al.* 2014)

There are scenarios where a manager could want to manage for biodiversity, but in many countries, laws such as the Endangered Species Act in the United States or Canada's Species at Risk Act (Congress 1973; Waples, Nammack, Cochrane, and Hutchings 2013), will require the manager or scientist to focus at the species level. The package presented in this article pretends to be a tool to change this either/or scenario and take information of both diversity and individual species models.

In this paper we present package **DiversityOccupancy**, used in the R environment, The objective of this package is to simultaneously model factors associated with occupancy and abundance of individual species using a detection history file, and to use predicted abundances to calculate species diversity for each sampling site. The package then models factor(s) associated with among-site species diversity, which can then be combined with spatial data to identify areas that contain both high abundance of species of conservation concern and high species diversity.

In the last decade, Occupancy modeling has been used more and more as a method to account for how species respond to environmental or anthropogenic factors. It has also been shown to be useful as a species distribution modeling tool when species have imperfect detection. Another use for what it has been used is for managers to change the environment of managed areas in order to improve the status of species of conservation concern (MacKenzie, Nichols, Lachman, Droege, Andrew Royle, and Langtimm 2002; MacKenzie 2006) but as far as we know this is the first method that takes into account both species diversity and individual species abundance in order to select conservation areas.

Occupancy models use detection-nondetection data from repeat surveys to simultaneously estimate probabilities of detection (p) and occupancy (ψ) (MacKenzie 2006). (Burnham

Site covariates Site covariates are presented in a data frame consisting of measurements taken at each site. The covariates are used singly and in combination to model occupancy or abundance, and they should be variables that are stable within the scope of the length of the study. In **DiversityOccupancy** there is an example concordant with the IslandBirds data set called siteCov:

```
data("siteCov")
head(siteCov, 10)
```

	Elev	AgroFo	SecVec	Wetland	Upland
1	214.6	0.04536772	0.309614772	0	0.6450175
2	254.7	0.00000000	0.013053168	0	0.9869468
3	321.5	0.00000000	0.009073543	0	0.9909265
4	68.2	0.00000000	0.000000000	0	1.0000000
5	346.5	0.00000000	0.106972302	0	0.8930277
6	74.5	0.00000000	0.000000000	0	1.0000000

Detection covariates A list of data frames, in which each data frame includes a daily measurement of variables with the potential to affect detection probabilities. It is important that each element (data frame) of the list has a name, so that it can be called to fit the occupancy model. These variables are used to model the probability of detection.

DiversityOccupancy has a data set called *Daily_Cov* which illustrates how the Daily covariates have to be structured:

```
#All the items of the list must have names
names(Daily_Cov)
```

```
[1] "Day"    "Wind"   "Obs"    "Time"   "Rain"   "Noise"  "Clouds"
```

```
#here we see the first dataframe of the Daily_Cov dataset
head(Daily_Cov[[1]])
```

	Day1	Day2	Day3	Day4
3	25	27	33	42
5	25	27	33	42
7	25	27	33	41
11	19	22	25	51
13	19	22	25	51
16	19	22	25	51

2.1. Fiting models for abundance and predicting alpha diversity

In this example we will fit and model the abundance for 5 bird species and calculate alpha diversity from those results.

```
birdDiversity <-diversityoccu(pres = IslandBirds, sitecov = siteCov,
obscov = Daily_Cov,spp = 5, form = ~ Day + Wind + Time ~ Elev + Wetland + Upland, dredge
```

The resulting object of class `diversityoccupancy` has the following elements

```
names(birdDiversity)
```

```
[1] "Covs"      "models"    "Diversity" "species"
```

If you need to see the parameters of the model of one of the species, you call the species number with the `element$models`. For example extract the model for the second species:

```
birdDiversity$models[[2]]
```

Call:

```
occuRN(formula = form, data = models[[i]])
```

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.66046	7.41e-01	0.891	0.3731
Elev	-0.00398	1.73e-03	-2.305	0.0212
Wetland	-47.31278	1.72e+02	-0.276	0.7828
Upland	-0.42087	5.56e-01	-0.756	0.4494

Detection:

	Estimate	SE	z	P(> z)
(Intercept)	-2.36626	1.80923	-1.308	0.1909
Day	0.03480	0.01801	1.932	0.0534
Wind	-0.10959	0.19166	-0.572	0.5675
Time	-0.00114	0.00375	-0.305	0.7605

AIC: 208.1134

The species parameter for a `diversityoccupancy` object shows us a table with the abundance and alpha diversity calculated for each sampled point:

```
summary(birdDiversity$species)
```

h	species.1	species.2	species.3	species.4
Min. :0.9332	Min. :0.1863	Min. :0.0000	Min. :1.636	Min. :0.1107
1st Qu.:1.2299	1st Qu.:0.4451	1st Qu.:0.1188	1st Qu.:2.128	1st Qu.:0.7577
Median :1.2952	Median :0.6292	Median :0.5799	Median :3.051	Median :2.4028
Mean :1.2679	Mean :0.6486	Mean :0.5679	Mean :2.740	Mean :1.9794
3rd Qu.:1.3262	3rd Qu.:0.7794	3rd Qu.:0.8910	3rd Qu.:3.275	3rd Qu.:3.0071
Max. :1.3616	Max. :1.2106	Max. :1.8616	Max. :3.553	Max. :3.7061

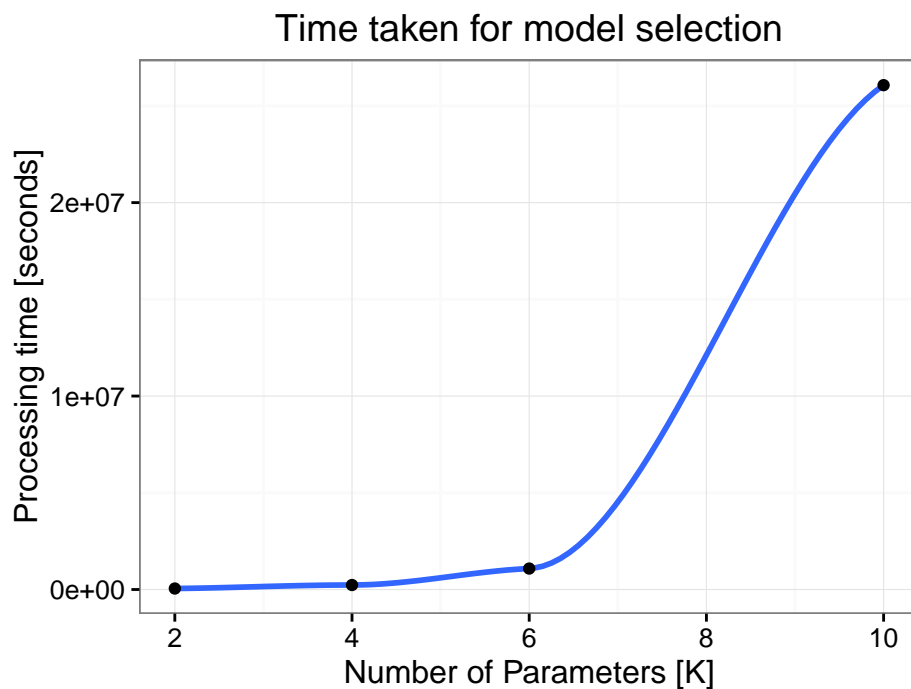
```

species.5
Min.    :0.04913
1st Qu.:0.29879
Median  :0.56381
Mean    :0.60893
3rd Qu.:0.74995
Max.    :2.01430

```

2.2. Automatic model selection for abundance models

If the option of dredge is set to “TRUE”, then `diversityoccu` attempts to fit all first order models, and it selects the one with the lowest AICc value, for each species. Be aware that processing times rapidly increases with added numbers of parameters, and that processing can require many hours or days for complex data sets. The following graph and table shows the processing time for the `IslandBirds` data set.



From now on we will work with automatically selected models for bird abundance and diversity using an information theoretic approach (AICc).

```

birdmodel.selected <- diversityoccu(pres = IslandBirds, sitecov = siteCov,
obscoV = Daily_Cov,spp = 5, form = ~ Day + Wind + Time ~ Elev + Wetland + Upland, dredge

```

Below we present an example of an analysis with the full model (includes all variables) and subsequently results from a model selection analysis, both of them only for the second species:

```

birdDiversity$models[[2]]

```

```
Call:
occuRN(formula = form, data = models[[i]])
```

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.66046	7.41e-01	0.891	0.3731
Elev	-0.00398	1.73e-03	-2.305	0.0212
Wetland	-47.31278	1.72e+02	-0.276	0.7828
Upland	-0.42087	5.56e-01	-0.756	0.4494

Detection:

	Estimate	SE	z	P(> z)
(Intercept)	-2.36626	1.80923	-1.308	0.1909
Day	0.03480	0.01801	1.932	0.0534
Wind	-0.10959	0.19166	-0.572	0.5675
Time	-0.00114	0.00375	-0.305	0.7605

AIC: 208.1134

```
birdmodel.selected$models[[2]]
```

```
Call:
occuRN(formula = ~Day + 1 ~ Elev + Wetland + 1, data = data2)
```

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.50953	0.66082	0.771	0.44067
Elev	-0.00446	0.00164	-2.725	0.00643
Wetland	-5.05318	2.71875	-1.859	0.06308

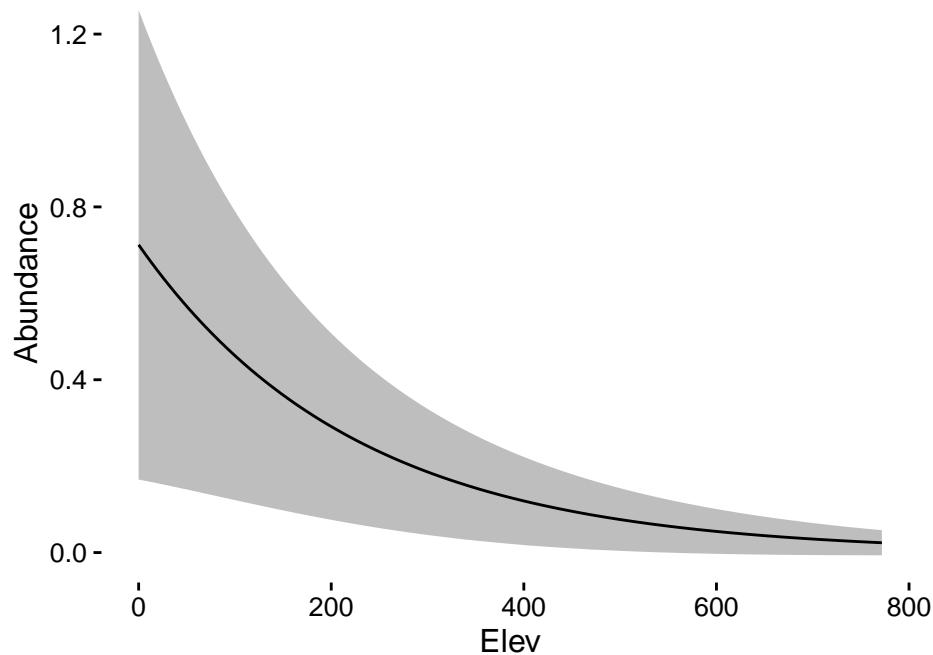
Detection:

	Estimate	SE	z	P(> z)
(Intercept)	-3.1681	0.8008	-3.96	7.62e-05
Day	0.0324	0.0174	1.86	6.29e-02

AIC: 204.0382

The responses of individual species to specific variables can be shown using the function `responseplot.abund`, below we show the response of abundance in species 2 to the Burn intensity soil. Note that this function automatically bounds the limits of the variable to the maximum and minimum observable values in the field.

```
responseplot.abund(birdmodel.selected, spp = 2, variable = Elev)
```



2.3. Model selection for alpha diversity modeling

The next step is to select the best model for predicting alpha diversity using the `model.diversity` function. The function takes a `diversityoccupancy` object, and fits all possible glm models and ranks them by AICc. Other than the `diversityoccupancy` object, there are three other parameters to select: 1) `Method`, which can be either “h” which fits every possible model, or “g”, which uses genetic algorithms to select models (recommended for large candidate sets); 2) `Delta`, which allows the user to identify an AICc delta threshold, which returns all models with AICc values below the threshold; 3) `Squared`, which includes only linear combinations when set to `FALSE` (Default), and both linear and quadratic (second order) if set to `TRUE`.

```
glm.BirdDiverse <- model.diversity(birdmodel.selected, method = "g", delta = 2,
squared = TRUE)
```

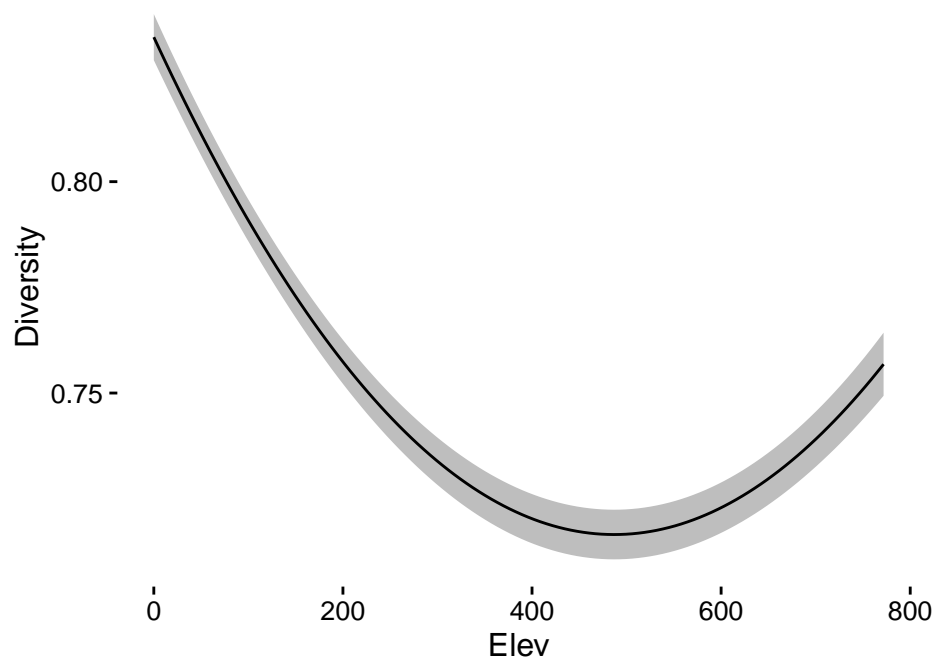
To see the top models extract the `Table` element of the `modelselection` object

```
\begin{CodeChunk} \begin{CodeOutput} model 1 Diversity ~ 1 + Elev + AgroFo + Wetland
+ I(Elev^2) + I(AgroFo^2) + I(SecVec^2) + I(Wetland^2) + I(Upland^2) 2 Diversity ~ 1 +
Elev + AgroFo + Wetland + I(Elev^2) + I(AgroFo^2) + I(Wetland^2) + I(Upland^2) 3 Diver-
sity ~ 1 + Elev + AgroFo + Wetland + Upland + I(Elev^2) + I(AgroFo^2) + I(Wetland^2) +
I(Upland^2) 4 Diversity ~ 1 + Elev + SecVec + Wetland + Upland + I(Elev^2) + I(AgroFo^2)
+ I(Wetland^2) + I(Upland^2) 5 Diversity ~ 1 + Elev + AgroFo + SecVec + Wetland +
I(Elev^2) + I(AgroFo^2) + I(Wetland^2) + I(Upland^2) aicc weights Delta.AICc 1 -681.3992
0.2610582 0.0000000 2 -681.1351 0.2287691 0.2640605 3 -680.7723 0.1908161 0.6268668 4 -
680.4609 0.1632999 0.9383094 5 -680.3701 0.1560568 1.0290469 \end{CodeOutput} \end{CodeChunk}
```

The `responseplot.diver` function takes a `modeldiversity` object and one of the variables used to predict the alpha diversity index, and makes a plot showing the response of the diversity index

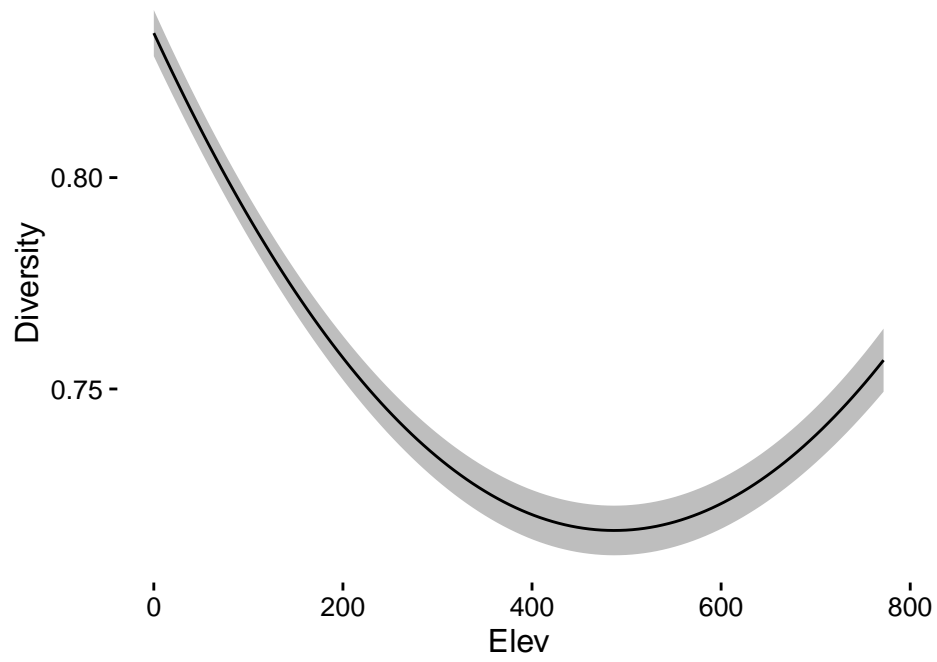
against the selected variable. This function automatically limits the values of that variable to the maximum and minimum values of the dataset. It also shows the standard deviation of the estimation.

```
responseplot.diver(glm.BirdDiverse, Elev)
```

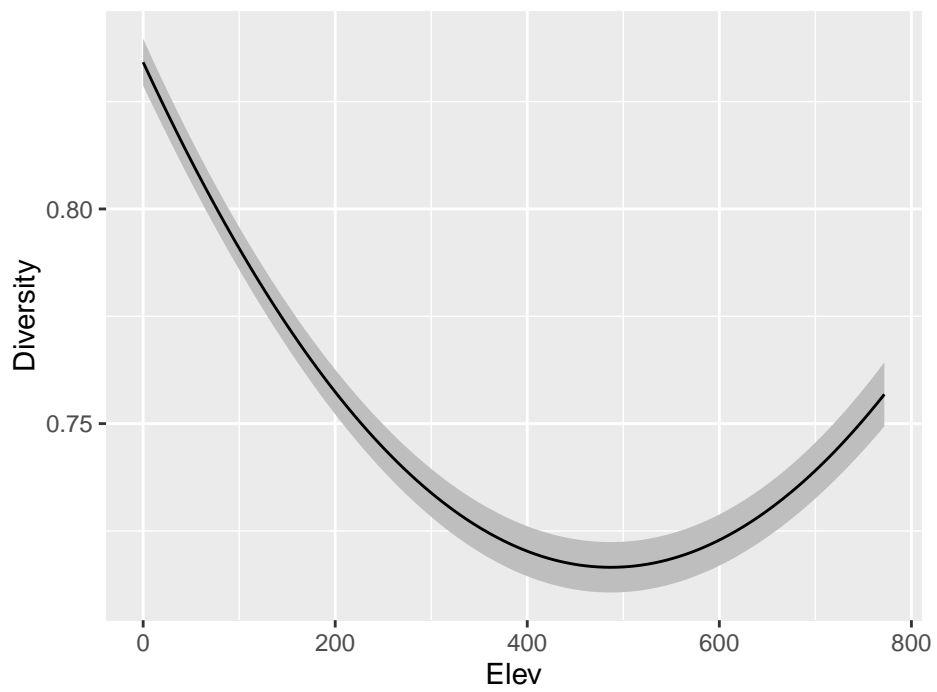


Also since the returned plot is a ggplot type object, it can be easily modified following ggplot2 grammar of graphics.

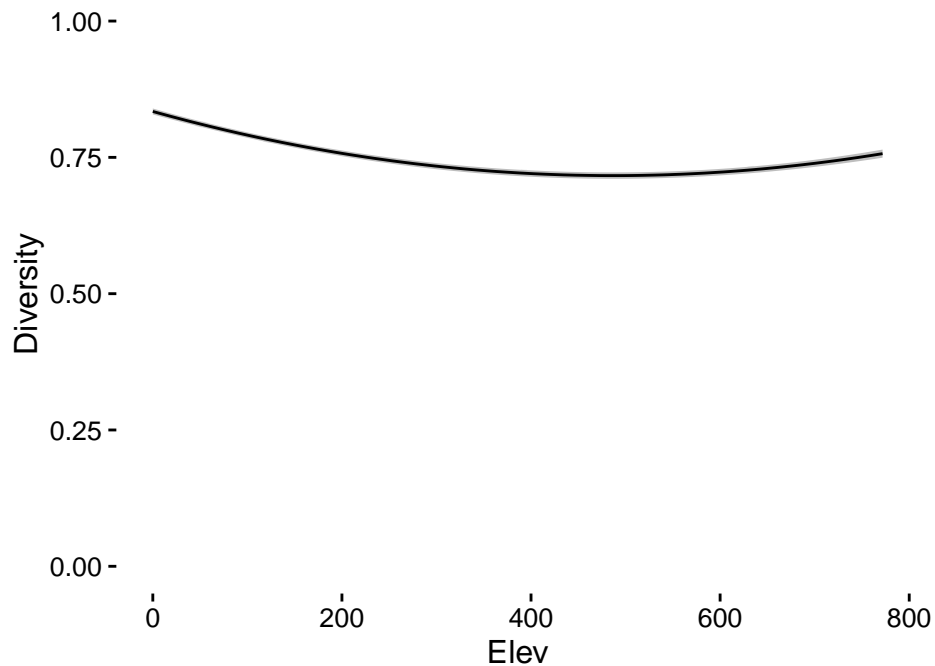
```
library(ggplot2)
k <- responseplot.diver(glm.BirdDiverse, Elev)
k
```



```
k + theme_gray()
```



```
k + ylim(c(0,1))
```



2.4. Selecting conservation areas based on alpha diversity and abundance of species of conservation concern

Of the 5 species modeled, let's say that there are at least two of conservation concern, the second and thirds of our list. Since we already have models relating site characteristics to species abundance and a model relating site characteristics to alpha diversity, with the use of a spatial raster layers (rasterstack) with the modeled variables we can choose an area with high species diversity and/or abundance. In order for this function to work properly the stack has to be in lon lat projection, and Birdstack is in UTM, so we reproject the data.

```
library(raster)
```

```
Loading required package: sp
```

```
Attaching package: 'sp'
```

```
The following object is masked from 'package:unmarked':
```

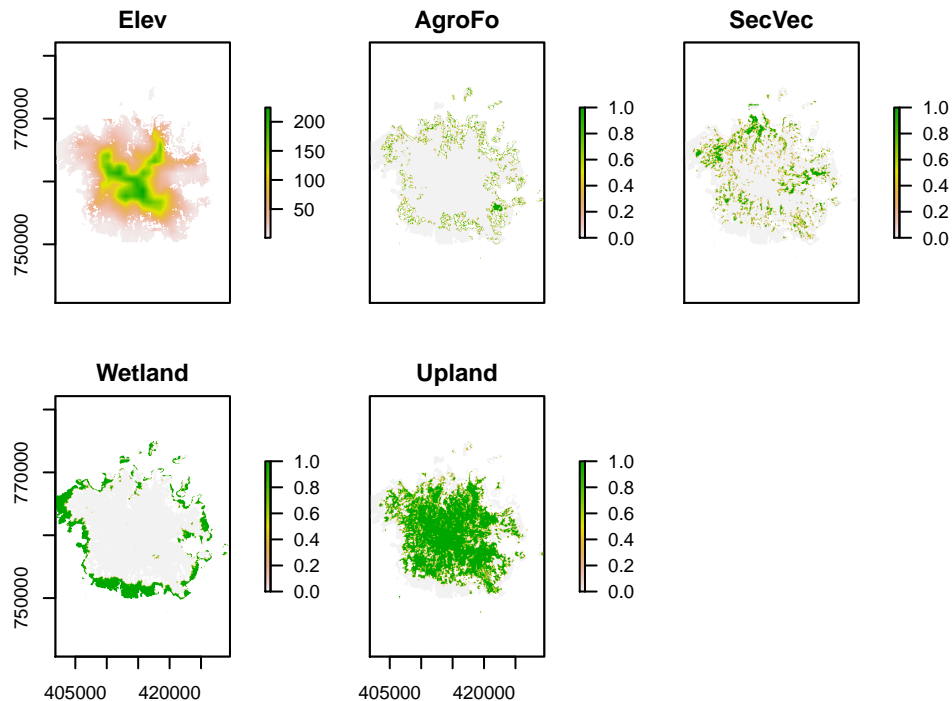
```
coordinates
```

```
Attaching package: 'raster'
```

```
The following objects are masked from 'package:unmarked':
```

```
getData, projection
```

```
data(Birdstack)
plot(Birdstack)
```



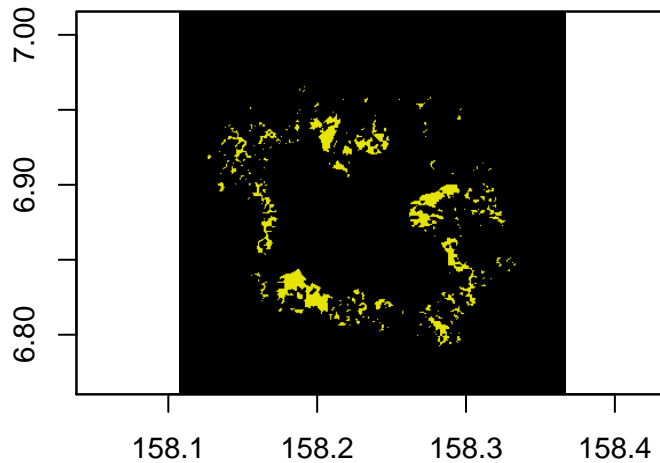
```
newproj <- '+proj=longlat +datum=WGS84 +ellps=WGS84 +towgs84=0,0,0'
library(rgdal)
```

```
rgdal: version: 1.1-9, (SVN revision 617M)
Geospatial Data Abstraction Library extensions to R successfully loaded
Loaded GDAL runtime: GDAL 2.0.1, released 2015/09/15
Path to GDAL shared files: C:/Users/corcoranbarriosd/Documents/R/win-library/3.2/rgdal/gdal
GDAL does not use iconv for recoding strings.
Loaded PROJ.4 runtime: Rel. 4.9.1, 04 March 2015, [PJ_VERSION: 491]
Path to PROJ.4 shared files: C:/Users/corcoranbarriosd/Documents/R/win-library/3.2/rgdal/proj
Linking to sp version: 1.2-3
```

```
Birdstack <- stack(projectRaster(Birdstack, crs=newproj))
```

In order to find areas of high conservation value, we use the `predict.diversity` function. We need both a `diversityoccupancy` and a `modeldiversity` class object, used in the model, and diverse parameters respectfully, a spatial representation of site covariates as raster files (`rasterstack`), with the variables in the `new.data` parameter, a boolean vector in the `species` parameter indicating which species shall be considered (T or F), and the `quantile.nth` parameter, which indicates a quantile threshold that is used for abundance and/or richness to indicate conservation value (areas above the threshold will be returned).

```
Selected.area <- diversity.predict(model = birdmodel.selected, diverse = glm.BirdDiverse,
c(F,T,T,F,F))
```



From the object `Selected.area` we can extract and plot not only the area that was selected with the attributes, but also the layers that show the predicted diversity for the island and also the predicted abundance for each of the species that were used for the analysis.

3. Discussion

References

- Bichet OM, Dupuch A, Hébert C, Le Borgne H, Fortin D (2016). “Maintaining animal assemblages through single-species management: the case of threatened caribou in boreal forest.” *Ecological Applications*.
- Branton M, Richardson JS (2011). “Assessing the Value of the Umbrella-Species Concept for Conservation Planning with Meta-Analysis.” *Conservation Biology*, **25**(1), 9–20.
- Burnham KP, Anderson DR (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.
- Congress U (1973). “Endangered species act.” *Washington DC*.
- Crosby AD, Elmore RD, Leslie DM, Will RE (2015). “Looking beyond rare species as umbrella species: Northern Bobwhites (*Colinus virginianus*) and conservation of grassland and shrubland birds.” *Biological Conservation*, **186**, 233–240.

- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, *et al.* (2013). "Predicting species distributions for conservation decisions." *Ecology letters*, **16**(12), 1424–1435.
- Keller V, Bollmann K (2004). "From red lists to species of conservation concern." *Conservation Biology*, **18**(6), 1636–1644.
- MacKenzie DI (2006). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press.
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Andrew Royle J, Langtimm CA (2002). "Estimating site occupancy rates when detection probabilities are less than one." *Ecology*, **83**(8), 2248–2255. URL [http://www.esajournals.org/doi/abs/10.1890/0012-9658\(2002\)083%5B2248:ESORWD%5D2.0.CO%3B2](http://www.esajournals.org/doi/abs/10.1890/0012-9658(2002)083%5B2248:ESORWD%5D2.0.CO%3B2).
- Möllmann C, Lindegren M, Blenckner T, Bergström L, Casini M, Diekmann R, Flinkman J, Müller-Karulis B, Neuenfeldt S, Schmidt JO, *et al.* (2014). "Implementing ecosystem-based fisheries management: from single-species to integrated ecosystem assessment and advice for Baltic Sea fish stocks." *ICES Journal of Marine Science: Journal du Conseil*, **71**(5), 1187–1197.
- Peterson AT (2011). *Ecological niches and geographic distributions (MPB-49)*. 49. Princeton University Press.
- Plagányi ÉE, Punt AE, Hillary R, Morello EB, Thébaud O, Hutton T, Pillans RD, Thorson JT, Fulton EA, Smith AD, *et al.* (2014). "Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity." *Fish and Fisheries*, **15**(1), 1–22.
- Pulliam HR (2000). "On the relationship between niche and distribution." *Ecology letters*, **3**(4), 349–361.
- Reza MIH, Abdullah SA, Nor SBM, Ismail MH (2013). "Integrating GIS and expert judgment in a multi-criteria analysis to map and develop a habitat suitability index: A case study of large mammals on the Malayan Peninsula." *Ecological indicators*, **34**, 149–158.
- ROBERGE JM, Angelstam P (2004). "Usefulness of the umbrella species concept as a conservation tool." *Conservation Biology*, **18**(1), 76–85.
- Rodrigues AS, Pilgrim JD, Lamoreux JF, Hoffmann M, Brooks TM (2006). "The value of the IUCN Red List for conservation." *Trends in ecology & evolution*, **21**(2), 71–76.
- Simberloff D (1998). "Flagships, umbrellas, and keystones: is single-species management passé in the landscape era?" *Biological conservation*, **83**(3), 247–257.
- Soniat TM, Conzelmann CP, Byrd JD, Roszell DP, Bridevaux JL, Suir KJ, Colley SB (2013). "Predicting the effects of proposed Mississippi River diversions on oyster habitat quality; application of an oyster habitat suitability index model." *Journal of Shellfish Research*, **32**(3), 629–638.

- Taft OW, Colwell MA, Isola CR, Safran RJ (2002). “Waterbird responses to experimental drawdown: implications for the multispecies management of wetland mosaics.” *Journal of Applied Ecology*, **39**(6), 987–1001.
- Tori GM, McLeod S, McKnight K, Moorman T, Reid FA (2002). “Wetland conservation and Ducks Unlimited: Real world approaches to multispecies management.” *Waterbirds*, pp. 115–121.
- Waples RS, Nammack M, Cochrane JF, Hutchings JA (2013). “A tale of two acts: endangered species listing practices in Canada and the United States.” *BioScience*, **63**(9), 723–734.
- White AM, Zipkin EF, Manley PN, Schlesinger MD (2013). “Conservation of avian diversity in the Sierra Nevada: moving beyond a single-species management focus.” *PloS one*, **8**(5), e63088.
- Zohmann M, Pennerstorfer J, Nopp-Mayr U (2013). “Modelling habitat suitability for alpine rock ptarmigan (*Lagopus muta helvetica*) combining object-based classification of IKONOS imagery and Habitat Suitability Index modelling.” *Ecological modelling*, **254**, 22–32.

Affiliation:

Derek Corcoran
University of Missouri
First line Second line
E-mail: corcoranbarriosd@missouri.edu
URL: <http://rstudio.com>