



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

Derek Corcoran
University of Missouri

Elisabeth Webb
University of Missouri, USGS

Dylan Kesler

Institute Of Bird Population

Abstract

Lately occupancy modeling has been vastly used as a tool for ecological research and management planing. However mostly it is used by interpreting single species models. We present the **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 then 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 usually 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 made to safe-ward the ones that are envisioned as species of conservation concern (Keller and Bollmann 2004; Rodrigues, Pilgrim, Lamoreux, Hoffmann, and Brooks 2006). Single species are easier to manage for, and it is easier to keep track the status of a species than it is to define management for diversity (Simberloff 1998), it is also more complicated to sample and keep track of changes in diversity.

There are several approaches to design 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 and Anderson 2003). We standardized all continuous site covariates t

1.2. Installing DiversityOccupancy

Requirements

To use this package you need R version 3.2.2 or newer (use the function `sessionInfo()` in your R session to check your current version).

Installing the package

Install from cran repository

```
install.packages("DiversityOccupancy")
```

1.3. Objectives of the Package

2. Use of the package

In order to calculate abundance and alpha diversity we need at least three files:

Detection history of multiple species A data frame consisting on the detection history of at least two species. As an example **DiversityOccupancy** has the data-set `BatOccu` which contains detection histories of 17 species of bats in the Plumas National Forest for 3 consecutive days (Columns) in 49 different sites (Rows). The data set includes a 1 for each time a species was detected, and a 0 for each time it was not detected.

A detection for the first three species is presented below:

```
library(DiversityOccupancy)
data("BatOccu")
head(BatOccu[1:9])
```

	Myyu1	Myyu2	Myyu3	Myca1	Myca2	Myca3	Myci1	Myci2	Myci3
1	0	0	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0	0	0
3	0	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	1	1	1	0	0	0
6	1	0	0	1	1	1	0	0	0

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 `BatOccu` data set called `sampling.cov`:

```
data("sampling.cov")
head(sampling.cov)
```

	Distance.to.water	Distance.to.road	Existing.vegetation	Fire.Interval	Altitude
1	0	325.2647	3.000000	14.79164	1859.337
2	0	0.0000	15.294588	11.00000	1839.813
3	0	0.0000	4.769200	16.00000	1890.586
4	0	0.0000	4.705464	18.27010	1927.237
5	0	0.0000	14.224747	14.97247	1682.559
6	0	2308.6010	15.727460	15.81841	1515.009

	Burn.intensity.soil	Burn.intensity.Canopy	Burn.intensity.basal
1	0.00000000	0.00000000	0.00000000
2	0.24802029	0.12812701	0.12812701
3	0.00000000	0.00000000	0.00000000
4	0.00000000	0.00000000	0.00000000
5	3.42075635	3.84151252	5.30799686
6	0.01135227	0.01135223	0.01135223

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 *Dailycov* which illustrates how the Daily covariates have to be structured:

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

```
[1] "Julian"    "Maxhum"    "Maxtemp"   "Meanhum"   "Meantemp"  "Minhum"    "Mintemp"
[8] "sdhum"     "sdtemp"
```

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

	Julian.Julian1	Julian.Julian2	Julian.Julian3
1	-1.683391	-1.683391	-1.683019
2	-1.620723	-1.620723	-1.620362
3	-1.684443	-1.684443	-1.684071
4	-1.557310	-1.557310	-1.556958
5	-1.429475	-1.429475	-1.434405
6	-1.241253	-1.241253	-1.240951

2.1. Fiting models for abundance and predicting alpha diversity

In this example we will fit and model the abundance for 17 bat species and calculate alpha diversity from those results.

```
\begin{CodeChunk} \begin{CodeInput} BatDiversity <-diversityoccu(pres = BatOccu, site-
cov = sampling.cov, obscov = Dailycov,spp = 17, form = ~ Julian + Meanhum ~ Burn.intensity.soil
+ I(Burn.intensity.soil^2), dredge = FALSE) \end{CodeInput} \end{CodeChunk}
```

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

```
names(BatDiversity)
```

```
[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:

```
BatDiversity$models[[2]]
```

Call:

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

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.000567	0.2829	0.002	0.998
Burn.intensity.soil	0.543374	0.3826	1.420	0.156
I(Burn.intensity.soil^2)	-0.092936	0.0996	-0.933	0.351

Detection:

	Estimate	SE	z	P(> z)
(Intercept)	0.113	0.357	0.317	0.7512
Julian	-0.097	0.267	-0.364	0.7159
Meanhum	-0.548	0.246	-2.228	0.0259

AIC: 180.113

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

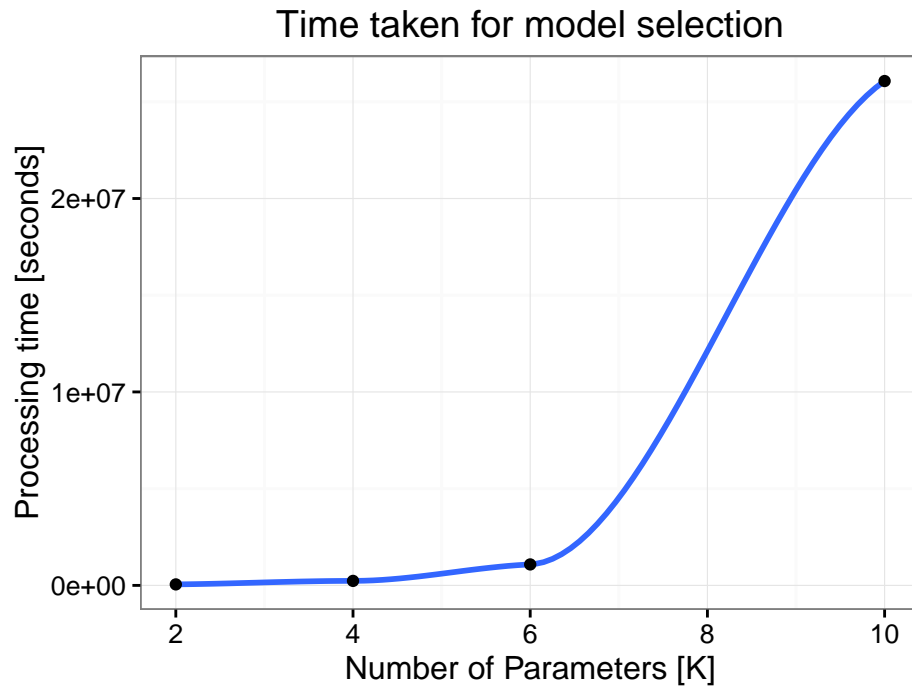
```
summary(BatDiversity$species)
```

h	species.1	species.2	species.3	species.4
Min. :2.115	Min. :0.3528	Min. :1.001	Min. :0.1156	Min. :0.4802
1st Qu.:2.115	1st Qu.:0.3528	1st Qu.:1.001	1st Qu.:0.1156	1st Qu.:0.5504
Median :2.277	Median :0.5153	Median :1.300	Median :0.1883	Median :0.5863
Mean :2.291	Mean :0.7520	Mean :1.523	Mean :0.3401	Mean :0.6125
3rd Qu.:2.488	3rd Qu.:1.2428	3rd Qu.:2.045	3rd Qu.:0.6356	3rd Qu.:0.5870
Max. :2.553	Max. :1.3027	Max. :2.214	Max. :0.6695	Max. :0.8884
species.5	species.6	species.7	species.8	species.9
Min. :0.3020	Min. :0.0000223	Min. : 2.776	Min. :1.716	Min. :0.1006

1st Qu.:0.3020	1st Qu.:0.0000223	1st Qu.: 4.261	1st Qu.:1.716	1st Qu.:0.2089
Median :0.5035	Median :0.0004349	Median : 6.424	Median :2.035	Median :0.3037
Mean :0.6907	Mean :0.0983938	Mean : 6.267	Mean :2.603	Mean :0.4328
3rd Qu.:1.1210	3rd Qu.:0.1874298	3rd Qu.: 6.424	3rd Qu.:3.584	3rd Qu.:0.3037
Max. :1.2776	Max. :0.3654810	Max. :12.836	Max. :4.150	Max. :1.6148
species.10	species.11	species.12	species.13	species.14
Min. :0.3994	Min. :0.5758	Min. :0.3391	Min. :1.638	Min. :1.016
1st Qu.:0.5334	1st Qu.:0.5770	1st Qu.:0.3391	1st Qu.:1.638	1st Qu.:1.239
Median :0.5334	Median :0.5816	Median :0.5080	Median :1.750	Median :1.360
Mean :1.5900	Mean :0.8988	Mean :0.5584	Mean :2.746	Mean :1.472
3rd Qu.:1.3059	3rd Qu.:1.0866	3rd Qu.:0.7826	3rd Qu.:3.533	3rd Qu.:1.369
Max. :7.1212	Max. :1.9802	Max. :0.9358	Max. :6.038	Max. :2.525
species.15	species.16	species.17		
Min. :0.1267	Min. : 0.7016	Min. :0.06252		
1st Qu.:0.1267	1st Qu.: 0.7016	1st Qu.:0.06252		
Median :0.2016	Median : 0.7636	Median :0.12145		
Mean :0.2653	Mean : 2.9118	Mean :0.36548		
3rd Qu.:0.4187	3rd Qu.: 3.3637	3rd Qu.:0.74717		
Max. :0.4708	Max. :11.9074	Max. :0.90081		

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 BatOccu data set.



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

```
\begin{CodeChunk} \begin{CodeInput} batmodel.selected <- diversityoccu(pres = BatOccu,
sitecov = sampling.cov, obscov = Dailycov, spp = 17, form = ~ Julian + Meanhum ~ Burn.intensity.soil
+ I(Burn.intensity.soil^2), dredge = TRUE) \end{CodeInput} \end{CodeChunk}
```

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:

```
BatDiversity$models[[2]]
```

Call:

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

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.000567	0.2829	0.002	0.998
Burn.intensity.soil	0.543374	0.3826	1.420	0.156
I(Burn.intensity.soil^2)	-0.092936	0.0996	-0.933	0.351

Detection:

	Estimate	SE	z	P(> z)
(Intercept)	0.113	0.357	0.317	0.7512
Julian	-0.097	0.267	-0.364	0.7159
Meanhum	-0.548	0.246	-2.228	0.0259

AIC: 180.113

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

Call:

```
occuRN(formula = ~Meanhum + 1 ~ Burn.intensity.soil + 1, data = data2)
```

Abundance:

	Estimate	SE	z	P(> z)
(Intercept)	0.0767	0.2637	0.291	0.7712
Burn.intensity.soil	0.1901	0.0973	1.953	0.0508

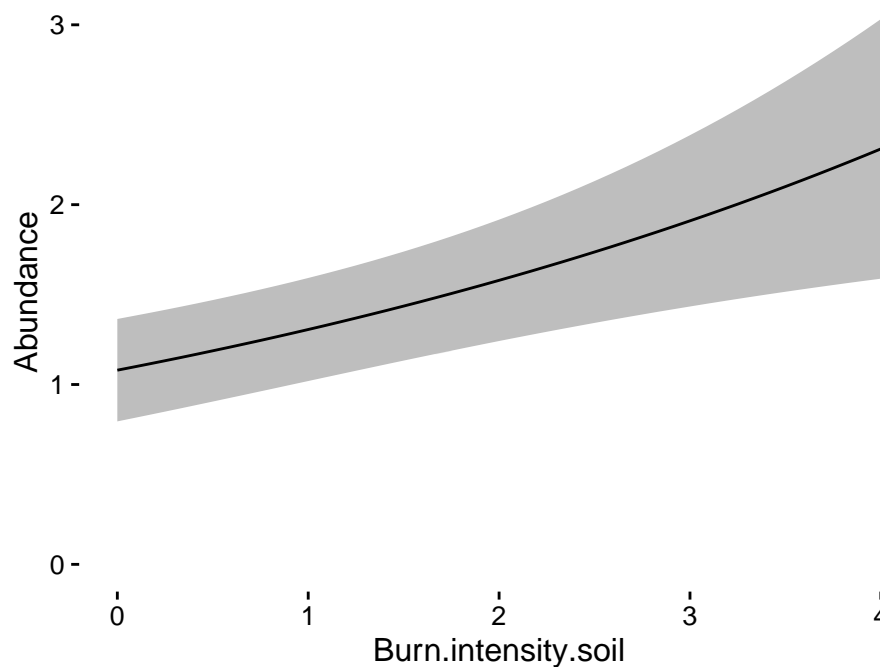
Detection:

	Estimate	SE	z	P(> z)
(Intercept)	0.143	0.351	0.407	0.6840
Meanhum	-0.530	0.242	-2.190	0.0285

AIC: 177.0765

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(batmodel.selected, spp = 2, variable = Burn.intensity.soil)
```



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.Batdiverse <- model.diversity(batmodel.selected, method = "g", delta = 2,
squared = TRUE)
```

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

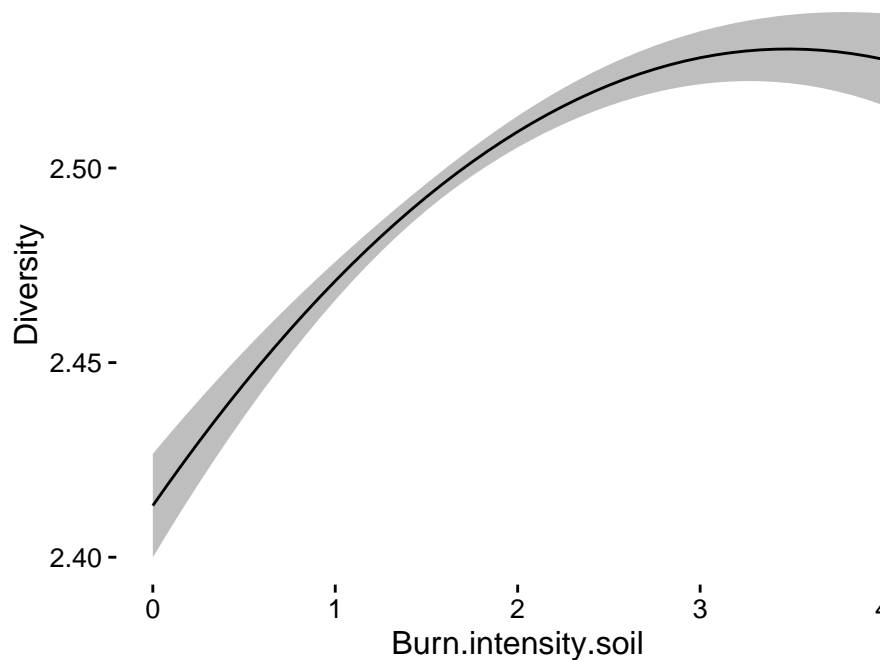
```
\begin{CodeChunk} \begin{CodeInput} class(glm.Batdiverse$Table) \end{CodeInput}
```

```
[1] "data.frame"
```

```
\end{CodeChunk}
```

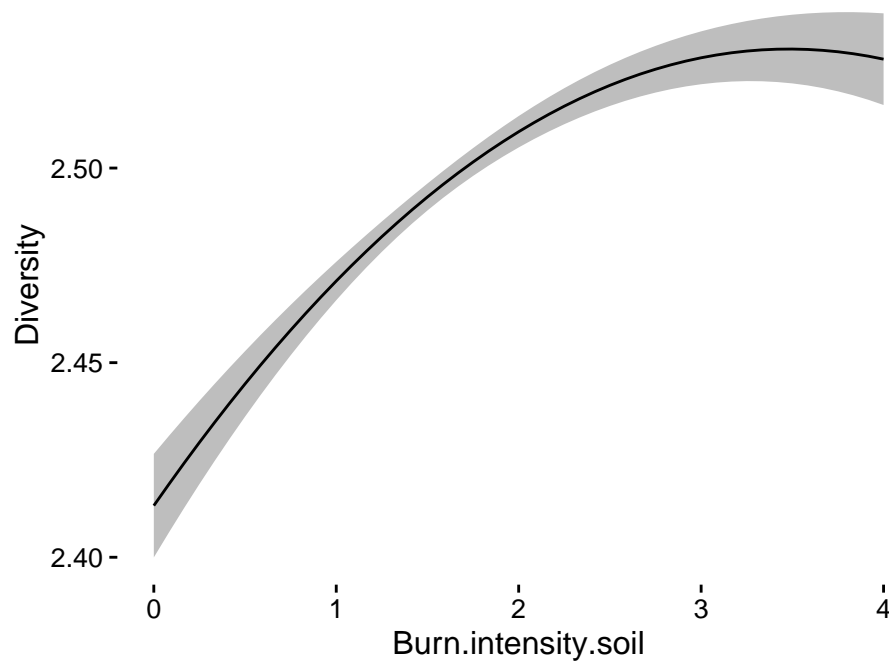
The `responseplot.diver` function takes a `modeldiversity` object and one of the variables used to predict the alpha diversity index, and makes a plot showing the response of the diversity index against the selected variable. This function automatically limits the values of that variable to the maximum and minimum values of the dataset. It also shows the standard deviation of the estimation.

```
responseplot.diver(glm.Batdiverse, Burn.intensity.soil)
```

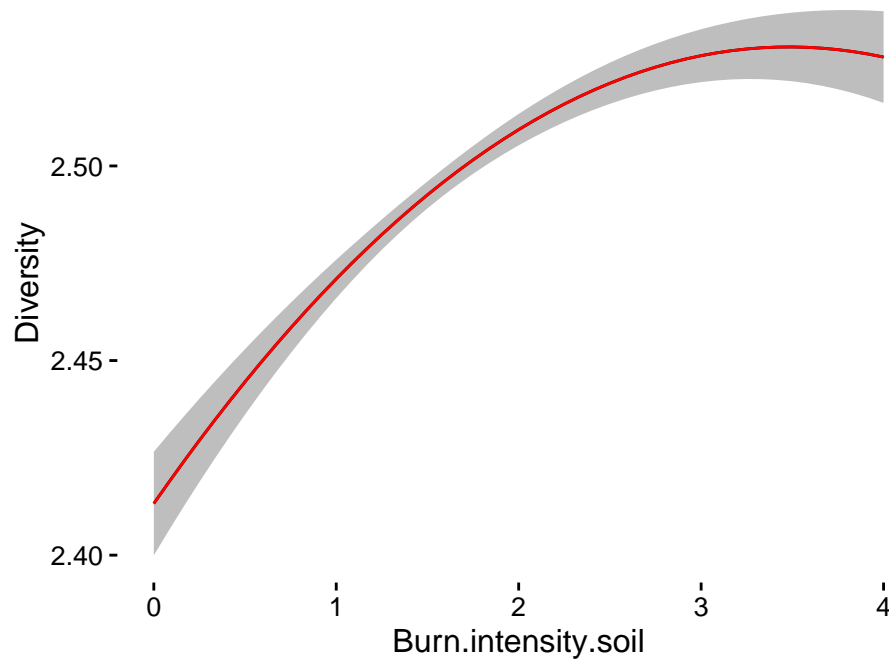


Also since the returned plot is a ggplot type object, it can be easily modified following ggplot2 grammar of graphics.

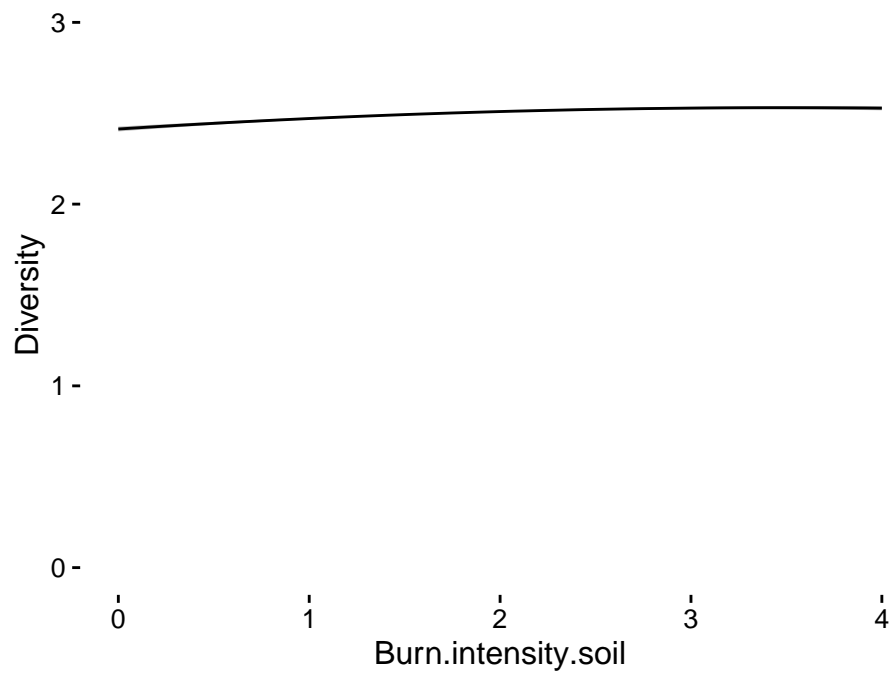
```
library(ggplot2)
k <- responseplot.diver(glm.Batdiverse, Burn.intensity.soil)
k
```



```
k + geom_line(color = "red")
```



`k + ylim(c(0,3))`



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

Of the 17 species modeled, there are at least three of conservation concern, pallid bat (*Antrozous pallidus*), corresponding to species 9th 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.

```
library(raster)
```

```
Loading required package: sp
```

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

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

```
coordinates
```

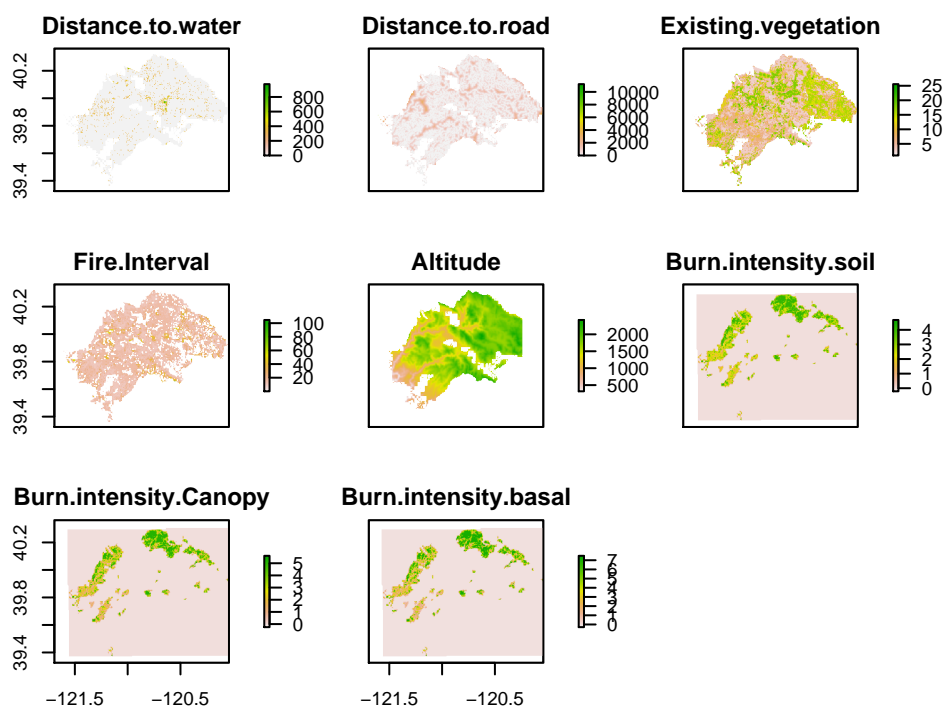
```
Attaching package: 'raster'
```

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

```
getData, projection
```

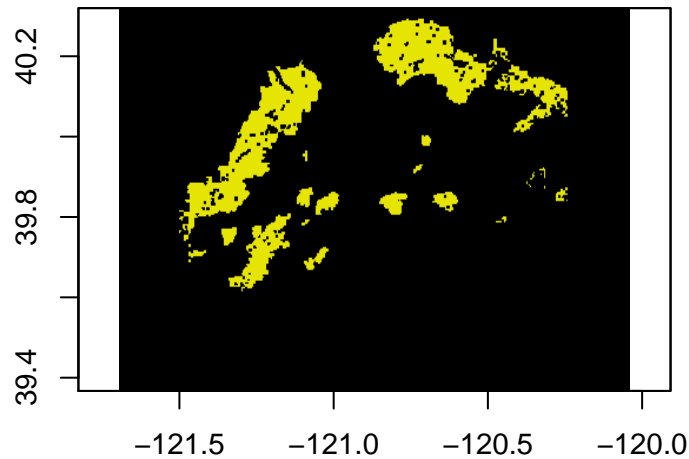
```
data(plumas.stack)
```

```
plot(plumas.stack)
```



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 = batmodel.selected, diverse = glm.Batdiverse, new.data =
c(F,F,F,F,F,F,F,F,T,F,F,F,F,F,F,F,F))
```



This function automatically creates a KMZ file that will be stored in the home directory of your session

3. Discussion

The **DiversityOccupancy** package lets scientists and managers take decisions based on species information, diversity information or both. In some countries, laws require that the decision is taken based on endangered species information, the possibility on selecting an area, or manage environments based on both diversity and species specific information, gives a possibility to managers or decision makers wanting to use diversity with laws requiring them to take species into account.

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>