

# rCAT: Species Conservation Assessment tools in R:

## Introduction and tutorials

Author: Justin Moat<sup>1,2</sup>

Contributors: Steve Bachman<sup>1,2</sup>

January 12, 2016

Document version: 1.1

1. Royal Botanic Gardens, Kew Richmond, Surrey, TW9 3AB, United Kingdom
2. School of Geography, Sir Clive Granger Building, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom

## Contents

|       |   |    |
|-------|---|----|
| 1     | Introduction .....  | 2  |
| 1.1   | Brief history .....   | 2  |
| 1.2   | Script development .....  | 2  |
| 1.3   | Discussion .....  | 2  |
| 1.3.1 | Cases of vagrancy, georeferencing and error checking .....                          | 2  |
| 1.3.2 | Extent of occurrence (EOO).....   | 3  |
| 1.3.3 | Area of occupancy (AOO).....  | 3  |
| 1.3.4 | Vulnerable D2 (VU D2) .....   | 3  |
| 1.3.5 | AOO and EOO thresholds.....   | 4  |
| 1.3.6 | <del>Number of localities and sub-populations (to be implemented later)</del> ..... | 4  |
| 1.3.7 | <del>Severely fragmented (to be implemented later)</del> .....                      | 4  |
| 2     | General rCAT usage .....  | 4  |
| 3     | Tutorial.....   | 5  |
| 3.1   | Installing rCAT package .....   | 5  |
| 3.2   | Quick single species example .....  | 5  |
| 3.3   | Batch processing .....  | 6  |
| 3.4   | Multiple species example (step through) .....                                       | 7  |
| 4     | Bibliography (essential reading) .....  | 10 |
| 5     | Additional references .....   | 10 |

# 1 Introduction

Red Lists are widely used to indicate species at risk of extinction. The tools presented here are scripts that integrate with primary occurrence data, to produce preliminary conservation assessments and to help assessors generate full species conservation assessments.

These tools are to be used with specimen data or sighting data with latitude and longitude co-ordinates and have been primarily designed to work with herbaria data which provides an important source of data relevant for Red List assessments.

The tools follow the IUCN Red Listing guidelines (IUCN 2012, 2016), and produce data for Extent of Occurrence (EOO), Area of Occupancy (AOO), ~~number of sub-populations (coming later)~~, ~~number of locations and indicators of fragmentation (in a later release)~~.

The conservation ratings given here, although calculated based on IUCN categories and criteria, are not equivalent to full IUCN ratings submissible for publication on the IUCN Red List. The majority of measures presented here relate to geometry and species range measures which focus on just one aspect of threat considered in the IUCN categories and criteria. Use of these tools assumes the user is familiar with the Red List Criteria (IUCN 2012, IUCN 2016).

However, the measures calculated do provide a good initial estimate of threat and can be used as a baseline for more detailed assessments that can incorporate population data, further literature, additional GIS analysis and consultation with experts and specialists.

These tools are primarily produced for quick assessment with the emphasis on speed of processing. Presently these tools do not produce any graphic or visual outputs. Use R's graphic outputs or <http://geocat.kew.org/> to visualise the data and investigate your species in the context of other mapping layers.

## 1.1 Brief history

These tools are formed and built on previous tools. A brief history of development is presented below for context and reference:

- 2003, first tool produced for ArcView 3.0 as used in Willis *et al.* 2003
- 2007, public release of cats.avx = Extension for ESRI ArcView 3.1 (Moat. 2007)
- 2011, release of GeoCAT online tools [geocat.kew.org](http://geocat.kew.org) (Bachman *et al* 2011)
- 2017, release of rCAT

## 1.2 Script development

Further improvements to the tool will be included in future releases. Note that any strikethrough text refers to functions that will be introduced at a later date.

## 1.3 Discussion

### 1.3.1 Cases of vagrancy, georeferencing and error checking

Care must be taken when using species spatial data in any analysis. Erroneous or incorrect outlying points or cases of vagrancy should be excluded and georeferencing should be error

checked prior to analysis. There are no data cleaning tools within rCAT, but many other tools and methods now exist to aid georeferencing and to describe errors for georeference uncertainty.

Specifically, in R see:

biogeo <https://cran.r-project.org/web/packages/biogeo/index.html> (Robertson 2016) and related paper: <http://onlinelibrary.wiley.com/doi/10.1111/ecog.02118/full>

Also see:

- See Wieczorek et al (2004) and the following links for details:
- <http://bg.berkeley.edu/latest/>
- <http://manisnet.org/GeorefGuide.html>
- <https://github.com/VertNet/api-geospatial/blob/master/GeospatialQuality.md>

Related paper:

<http://bioinformatics.oxfordjournals.org/content/early/2016/02/16/bioinformatics.btw057.short>

- <http://www.gbif.org/prog/digit/Georeferencing>
- <http://www.museum.tulane.edu/geolocate/default.aspx>
- <http://www.diva-gis.org/>

### 1.3.2 Extent of occurrence (EOO)

The IUCN Guidelines (Version 12 – February 2016) provides useful discussion of the problems associated with estimating EOO. The minimum convex polygon (MCP), as used here, has previously been criticised due to the coarse level of resolution on its outer surface (Ostro et al. 1999). However, the MCP is still considered a useful approximation of EOO as the purpose of EOO is to measure the degree to which risks from threatening factors are spread spatially across the taxon's geographical range (Joppa *et al.* 2015). When EOO is applied to criteria other than B1, such as reductions (criterion A) or continuing declines (B and C), then the MCP may no longer be the most appropriate measure as it is sensitive to outliers and sampling effort. An alpha hull may be more appropriate in this situation (Burgman and Fox, 2003).

### 1.3.3 Area of occupancy (AOO)

The problems associated with scale and AOO measurements are well described in the IUCN guidelines. The tools in this extension allow the user to define a grid cell length in metres. ~~In addition a sliding scale grid cell width is suggested which is scaled to the EOO measurement (Willis et al 2003).~~ The IUCN recommended scale of a 2 x 2 km grid (4 km<sup>2</sup>) can be applied but may not be appropriate if the point distribution does not fully represent occupied areas. In this case the assumption will be that areas without points are confirmed absences whereas this may simply be an artefact of collection effort. The approach described above is deemed suitable for a preliminary assessment, but for a full assessment it is advised that the AOO measure is more carefully defined e.g. by using a downscaling approach (Marsh 2016) or using habitat maps derived from remote sensing imagery or range prediction models. The IUCN guidelines document should be consulted when using habitat maps as estimates for AOO.

### 1.3.4 Vulnerable D2 (VU D2)

As well as the B criterion the AOO measure can also be applied to the D criterion under Vulnerable (VU) D2. If AOO is restricted such that the population is prone to the effects of human activities or stochastic events and may become Critically Endangered or Extinct in a

short time period then VU D2 may be appropriate. However, although a guideline threshold of 20 km<sup>2</sup> is suggested by IUCN this is not intended to be exclusive and therefore it is not reported in the preliminary IUCN results created by this tool.

### 1.3.5 AOO and EOO thresholds

We have used the thresholds described in the IUCN Red Listing guidelines (IUCN 2012, 2016). We have also adopted the Near Threatened threshold following the examples given in IUCN Standards and Petitions Subcommittee guidelines version 11 (2014). Summary of these thresholds and abbreviations are in the table below.

| IUCN Rating           | Rating abbreviation | AOO threshold (km <sup>2</sup> ) | EOO Threshold (km <sup>2</sup> ) |
|-----------------------|---------------------|----------------------------------|----------------------------------|
| CRITICALLY ENDANGERED | CR                  | < 10                             | < 100                            |
| ENDANGERED            | EN                  | < 500                            | < 5,000                          |
| VULNERABLE            | VU                  | < 2,000                          | < 20,000                         |
| NEAR THREATENED       | NT                  | < 3,000                          | < 30,000                         |

Table 1. IUCN category thresholds used rCAT

### ~~1.3.6 Number of localities and sub-populations (to be implemented later)~~

### ~~1.3.7 Severely fragmented (to be implemented later)~~

~~The AOO Density measure can be used as an indication of a fragmented distribution, although will not be enough to qualify a species under the IUCN 'severely fragmented' sub-criterion. In order to qualify, a detailed understanding of area of occupancy, dispersal ability and average population density is now required. It is hoped a new model for severely fragmented can be incorporated into future updates of this extension.~~

## 2 General rCAT usage

This is a relatively simple package, with several additional routines for data manipulation. The general work order is:

1. Import of species occurrence data in geographic projection e.g. latitudes and longitudes. If data is already projected to meters, skip point 2 below\*
2. Project geographic data to suitable area projection in metres. N.B. this can be either for all species, if for a localised area or individual species for more widespread species data. We use the geographic centre of the set of points to project the data.
3. Calculate EOO area and AOO area
4. Return IUCN categories for EOO and AOO areas

\* The assumption is that the data is cleaned, is in decimal degrees and represents wild occurrences of the taxon to be assessed. See 1.3.1 for discussion and list of tools to support data cleaning.

## 3 Tutorial

### 3.1 Installing rCAT package

To install the package from CRAN

```
install.packages("rCAT")
```

To load the library

```
library("rCAT")
```

### 3.2 Quick single species example

As a quick example, we'll import and produce metrics for a single species. The data we'll use is dummy data, in csv from [geocat.kew.org](http://geocat.kew.org). I have added this dummy data into this package to retrieve this data use `system.file("extdata","species_x.csv",package="rCAT")`. To use your own data replace this with your file (ie `my_data.csv`)

```
mydata <- read.csv(system.file("extdata","species_x.csv",package="rCAT"))
```

The `lat` (`mydata$latitude`) and `long` (`mydata$longitude`) fields are needed for any analysis.

```
mypointsll <- data.frame(lat=mydata$latitude,long=mydata$longitude)
```

To project the data automatically to an area projection we are going to use the projection wizard, firstly we'll need to get the true geographic centre of the points.

```
#find the true centre of the points
```

```
centreofpoints <- trueCOGll(mypointsll)
```

```
#project to an equal area projection
```

```
mypointsxy <- simProjWiz(mypointsll,centreofpoints)
```

We now have a set of points (`mypointsxy`) in an equal area projection in metres, ready to calculate EOO in  $\text{km}^2$ . Note area is calculated in  $\text{m}^2$ , so we need to convert to  $\text{km}^2$ . Also area is returned as negative (as polygon is constructed anticlockwise)

```
#Calculate the Extent of Occurrence EOO
```

```
EOOm2 <- EOOarea(mypointsxy)
```

```
EOOkm2 <- EOOm2/1000000
```

Now for the AOO in  $\text{km}^2$ . Note `AOOsimp` returns the number of cells not the area and `cellsize` is the length of one side of the square cell in metres

```
#Calculate the Area of Occupancy AOO for 2km cells
```

```
cellsize <- 2000
```

```
AOOnocells <- AOOsimp(mypointsxy,cellsize)
```

```
AOOkm2 <- AOOnocells * (cellsize/1000)^2
```

To return the IUCN category for these EOO and AOO values. Note the second parameter triggers return of the full text or the abbreviation for the IUCN category.

```
#review the IUCN category for this AOO and EOO Area
```

```
EOOcat <- EORating(EOOkm2,FALSE)
```

```
AOOcat <- AORating(AOOkm2,FALSE)
```

Results are in km<sup>2</sup> are below with the IUCN categories:

```
#show results
```

```
> EOOkm2
```

```
[1] 585915.6
```

```
> EOOcat
```

```
[1] "Least Concern"
```

```
> AOOkm2
```

```
[1] 56
```

```
> AOOcat
```

```
[1] "Endangered"
```

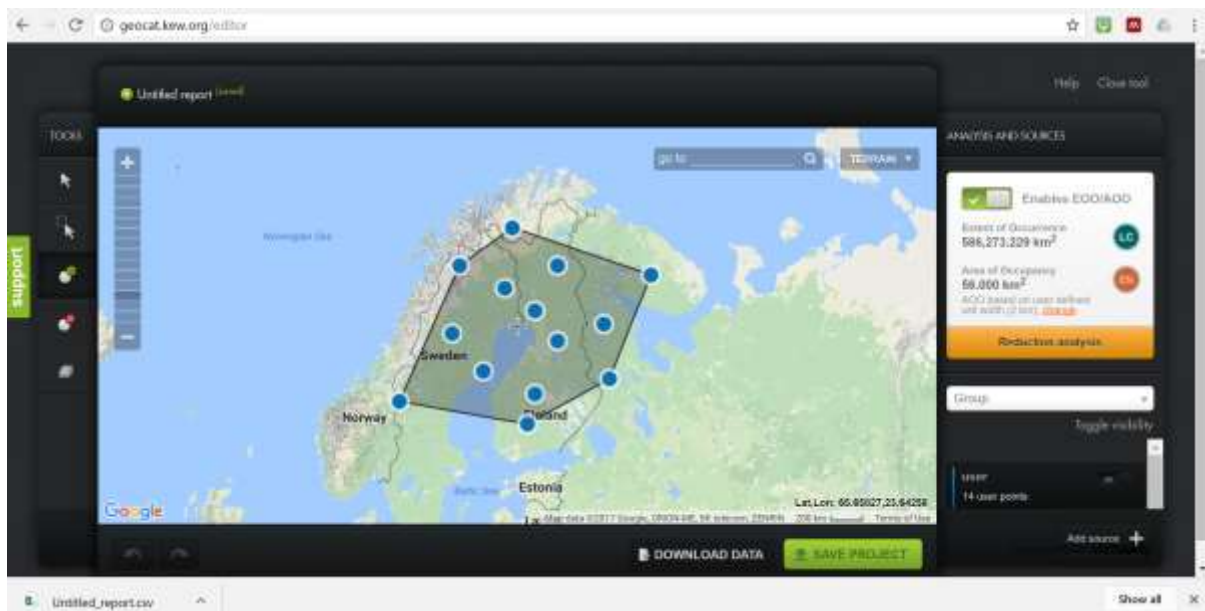


Figure 1. Results for species *x* in GeoCAT

### 3.3 Batch processing

The above example is for a single species, but the code is designed to perform with large numbers of species and across large databases. Here we will work with a dummy dataset of 4 species (w,x,y and z).

- Species w is a global (land and sea) species
- Species x as above single species example
- Species y is a Peru only species
- Species z is a North polar species

First step is to read the dataset and also get a unique list of species. I have supplied a dummy dataset for testing `system.file("extdata","multiple_species.csv",package="rCAT")`, but you can replace this with your dataset i.e. `myfile.csv`

This csv is formatted:

```
scientificname,latitude,longitude
species x,67.06743335,22.5
species x,65.36683689,17.578125
species x,62.59334083,12.65625
species x,62.91523304,25.3125
```

We will use the function `ConBatch` which combines all the routines into one simple function, returning a dataframe (table) or results by species

First step is to read the dataset. I have supplied a dummy dataset for testing `system.file("extdata","multiple_species.csv",package="rCAT")`, but you can replace this with your dataset i.e. `myfile.csv`

```
#read in simple csv data
mydata <- read.csv(system.file("extdata","multiple_species.csv",package="rCAT"))
```

Next is to call `ConBatch`, in this case we are using a cell length of 2000m (2km). The final argument is either `TRUE` or `FALSE`. `TRUE` is where all data is projected together using the centre of all latitudes and longitudes. `FALSE` each species is projected separately. I would suggest you use `TRUE` if data is all from a similar area (i.e. all from one country/region), as this will speed up processing.

```
results <- ConBatch(mydata$scientificname,mydata$latitude,mydata$longitude,2000,FALSE)
```

Final results are in dataframe `results`. If you wish to save this as a csv:

```
write.csv(results, file="results.csv")
```

|   | taxa      | NOP  | MER              | EOOkm2           | AOO2km | EOOcat | AOOcat |
|---|-----------|------|------------------|------------------|--------|--------|--------|
| 1 | species x | 14   | 918562.259811711 | 585915.607417865 | 14     | LC     | EN     |
| 2 | species y | 124  | 1717224.64389286 | 634149.482670821 | 124    | LC     | EN     |
| 3 | species z | 61   | 22622717.2314339 | 17839113.1220552 | 61     | LC     | EN     |
| 4 | species w | 1130 | 509390660.388499 | 506445176.073246 | 1130   | LC     | VU     |

Table 2. Example results from `ConBatch` for dummy dataset

### 3.4 Multiple species example (step through)

The above example is very simple, but not flexible. The functions in `rCAT` are atomised so you can call each individually. The example below, uses the same data as above, stepping through each conservation metric, but testing two cell sizes.

As above first step is to read the dataset and also get a unique list of species. I have supplied a dummy dataset for testing `system.file("extdata","multiple_species.csv",package="rCAT")`, but you can replace this with your dataset i.e. `myfile.csv`

```
#read in simple csv data

mydata <- read.csv(system.file("extdata","multiple_species.csv",package="rCAT"))

mypointsll <-
data.frame(taxa=mydata$scientificname,lat=mydata$latitude,long=mydata$longitude)

specieslist <- unique(mypointsll$taxa)
```

Next to build a data frame to store the results

```
#make a dataframe to store results
resultsdf<- data.frame(taxa=character(),NOP=integer(),
MER=double(),EOOkm2=double(),AOO2km=double(),AOO100km=double(),stringsAsFactors=F
ALSE)
```

We then just need to loop through each taxon, extracting one at a time.

```
for (taxa in specieslist){
print(paste("Processing",taxa))
taxapointsll <- (mypointsll[mypointsll$taxa==taxa,c("lat", "long")])
}
```

As for single species we'll re-project to an equal area projection and then calculate the various metrics we need. In the example script we calculate:

| Column heading | Description  |
|----------------|--|
| Taxa           | Species name   |
| NOP            | Number of points used for analysis                               |
| MER            | Area of the minimum enclosing rectangle in km <sup>2</sup>       |
| EOOkm2         | Area of EOO in km <sup>2</sup>                                   |
| AOO2km         | Number of cells for a 2 x 2 km cell (4 km <sup>2</sup> )         |
| AOO100km       | Number of cells for a 100 x 100 km cell (10000 km <sup>2</sup> ) |

Table 3. Column headings for results dataframe

Implementing the loop above, with all of the analysis (note loop repeated from above for consistency).

```
for (taxa in specieslist){
print(paste("Processing",taxa))

#extract one taxa

taxapointsll <- (mypointsll[mypointsll$taxa==taxa,c("lat", "long")])
```



```

#reproject points
taxapointsxy <- simProjWiz(taxapointsll,trueCOGll(taxapointsll))

#CALCULATE METRICS

#number of points
nop <- nrow(taxapointsxy)

#Minimum enclosing rectangle
MERps <- MER(taxapointsxy)/1000
MERarea <- (MERps[2] - MERps[1]) * (MERps[4]- MERps[3])

#EOO
EOOa <- EOOarea(taxapointsxy)/-1000000

#AOO with 2 x 2 km cells
AOO2km <- AOOsimp(taxapointsxy,2000)

#AOO with 100 x 100 km cell
AOO10km <- AOOsimp(taxapointsxy,100000)

#finally in this loop we need to populate the dataframe
#with the results for each taxon
resultsdf[nrow(resultsdf)+1,] <- c(taxa,nop,MERarea,EOOa,AOO2km,AOO10km)
} #end of loop

```

Final results are in the dataframe resultsdf:

|   | taxa      | NOP  | MER              | EOOkm2           | AOO2km | AOO100km |
|---|-----------|------|------------------|------------------|--------|----------|
| 1 | species x | 14   | 918562.259811711 | 585915.607417865 | 14     | 14       |
| 2 | species y | 124  | 1717224.64389286 | 634149.482670821 | 124    | 65       |
| 3 | species z | 61   | 22622717.2314339 | 17839113.1220552 | 61     | 61       |
| 4 | species w | 1130 | 509390660.388499 | 506445176.073246 | 1130   | 1111     |

Table 4. Example results for dummy datasets

If you wish to save the results as csv:

```
write.csv(resultsdf,file="results.csv")
```

## 4 Bibliography (essential reading)

Bachman, S., Moat, J., Hill, A. W., de Torre, J., & Scott, B. (2011). Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. *ZooKeys*, 126(150), 117–26. <https://doi.org/10.3897/zookeys.150.2109>

IUCN. (2012). *IUCN RED LIST CATEGORIES AND CRITERIA* (2nd ed.). IUCN, Gland, Switzerland. <http://www.iucnredlist.org/technical-documents/categories-and-criteria/2001-categories-criteria>

IUCN Standards and Petitions Subcommittee. (2016). *Guidelines for Using the IUCN Red List Categories and Criteria. Version 12*. Retrieved from <http://www.iucnredlist.org/documents/RedListGuidelines.pdf>

Willis, F., Moat, J., & Paton, A. (2003). Defining a role for herbarium data in Red List assessments : a case study of *Plectranthus* from eastern and southern tropical Africa. *Biodiversity & Conservation*, 12(7), 1537–1552. Retrieved from <http://www.springerlink.com/index/M7588R02L2585577.pdf>

## 5 Additional references

Burgman, M. and, Fox, J.C., 2003. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Anim. Conserv.* 6, 19–28. doi:10.1017/S1367943003003044

IUCN Standards and Petitions Subcommittee, 2014. Guidelines for Using the IUCN Red List Categories and Criteria. Version 11.

Joppa, L.N., Butchart, S.H.M., Hoffmann, M., Bachman, S.P., Akçakaya, H.R., Moat, J.F., Böhm, M., Holland, R.A., Newton, A., Polidoro, B., Hughes, A., 2016. Impact of alternative metrics on estimates of extent of occurrence for extinction risk assessment. *Conserv. Biol.* 30, 362–370. doi:10.1111/cobi.12591

Marsh Charlie. (2016). Package “downscale” 1.2-2. Retrieved from The package provides a set of functions that model the occupancy-area relationship (OAR) of known coarse scale data. The models are then extrapolated to predict the proportion of occupied area at finer grain sizes.

Moat, J. F. (2007). Conservation assessment tools extension for ArcView 3.x, version 1.2. Retrieved from <http://www.kew.org/gis/projects/cats/catsdoc.pdf>

Moat, J., & Bachman, S. (2017). GeoCAT Geospatial Conservation Assessment Tool. Retrieved from <http://geocat.kew.org/>

Ostro, L.E., Young, T.P., Silver, S.C., Koontz, F.W., 1999. A Geographic Information System Method for Estimating Home Range Size. *Source J. Wildl. Manag.* 63, 748–755.

Robertson, M.P., Visser, V., Hui, C., 2016. Biogeo: An R package for assessing and improving data quality of occurrence record datasets. *Ecography (Cop.)*. 39, 394–401. doi:10.1111/ecog.02118

Wieczorek, J., Guo, Q., Hijmans, R.J., n.d. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty.  
doi:10.1080/13658810412331280211