

Ecography

**E7138**

Cruse, B., Liedloff, A. C. and Wintle, B. A. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. – *Ecography* 35: xxx–xxx.

**Supplementary material**

# Appendix 1

In this appendix we provide code to enable users to implement the residuals autocovariate (RAC) model as a GLM or a boosted regression tree (BRT) using the statistical package R (version 2.11.1, R Core Development Team, 2010). Implementation of a simple model containing only environmental predictors, and the autologistic model are also demonstrated to enable users to modify the code for their own data, and to compare the three modelling approaches. The dataset, "Snouter" is available from <http://www.oikos.ekol.lu.se/appendixdown/snouterdata.txt>, and was originally supplied as supplementary material to Dormann et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. - *Ecography* 30: 609-628.

For more about implementing BRT see Elith, Leathwick and Hastie. 2008. A working guide to boosted regression trees. - *Journal of Animal Ecology* 77: 802-813. For details of the "raster" package, refer to Hijmans. 2011. The rasterfile format, available from [cran.r-project.org/web/packages/raster/vignettes/rasterfile.pdf](http://cran.r-project.org/web/packages/raster/vignettes/rasterfile.pdf). Here we use focal operations to derive the autocovariate term, however, inverse distance weighting (or other types of weighting schemes) can be applied, see Dormann et al. 2007. *Ecography* 30: 609-628.

```
#Set the working directory and read in your data

#Data here called "Snouter"

setwd("E:\\Example")

Snouter <- read.table(file="E:\\Example\\Snouter_data.txt", header=TRUE)

#Read in libraries

require(gbm) #for BRT models

source("brt.functions.R") #for BRT models

source("model_functions.R") #for BRT models

require(raster) #for focal calculations

xy <- cbind(Snouter$X, Snouter$Y) #extract xy co-ordinates from data

#####
```

```

## GLMs

#####

#Fit the model with environmental variables

env_glm <- glm(value1 ~ rain + djungle, Snouter, family = binomial)

#####

#RAC model (autocovariate derived from residuals of model with environmental predictors)

#Derive the autocovariate term from focal operation

#Set up a blank rasterfile

rast <- raster(ncol=39, nrow = 47, ymn = 1, ymx = 40, xmn = 1, xmx = 50)
res(rast) <- 1

#Extract residuals from the model called "env_glm" and map them
xy_residuals <- cbind(xy, resid(env_glm))
rast[cellFromXY(rast, xy_residuals)] <- xy_residuals[,3]

plot(rast)

#Calculate residuals autocovariate

#Focal operations: ngb is neighbourhood size, set to 3 by 3 cells; fun is function,
#here the mean value within the defined neighbourhood

focal_rac_rast <- focal(rast, ngb = 3, fun = mean, na.rm = TRUE)

plot(focal_rac_rast)

#Extract the values of the focal operation from "focal_rac_rast" rasterfile using the
#co-ordinates stored in "xy"

focal_rac_vect <- xyValues(focal_rac_rast, xy)

focal_rac_vect <- xyValues(focal_rac_rast,xy)

#Add as a column to the data

Snouter<- cbind(Snouter, focal_rac_vect)

# fit the RAC model using the environmental variables and the residuals autocovariate

```

```
rac_glm <- glm(value1 ~ rain + djungle + focal_rac_vect, Snouter, family = binomial)
```

```
#####
```

```
#Autologistic model
```

```
#Derive the autocovariate from the response variable (ie presence/absence of Snouter)
```

```
#Set up blank rasterfile
```

```
rast_ac <- raster(ncol=39, nrow = 47, ymn = 1, ymx = 40, xmn = 1, xmx = 50)
```

```
res(rast_ac) <- 1
```

```
  #fill the raster with the values of Snouter presence or absence allocated using "xy"
```

```
  rast_ac[cellFromXY(rast_ac, xy)] <- Snouter[,5]
```

```
  #calculate the autocovariate via a focal operation
```

```
  focal_response_rast <- focal(rast_ac, ngb = 3, fun = mean, na.rm = TRUE)
```

```
  plot(focal_response_rast)
```

```
  ac_vect <- xyValues(focal_response_rast, xy)
```

```
  Snouter <- cbind(Snouter, ac_vect)
```

```
  #fit the autologistic model
```

```
  autolog_glm <- glm(value1 ~ rain + djungle + ac_vect, Snouter, family = binomial)
```

```
#####
```

```
## BRT models
```

```
#####
```

```
#BRT model with environmental variables
```

```
env_brt <- gbm.step(data = Snouter,
```

```
gbm.x = 3:4,          #column number for predictor variables
```

```
gbm.y = 5,           #column number for response variable
```

```
family = "bernoulli", #bernoulli for presence/absence data
```

```
tree.complexity = 3, #number of nodes in the tree, 3 allows for some interactions
```

```
learning.rate = 0.002,          #slow this down to ensure at least 1000 tree are fitted
bag.fraction = 0.5)           #introduces stochasticity, 0.5 is the default setting
```

```
#####
```

```
#RAC model (residuals based autocovariate)
```

```
#First map the residuals from the env_brt model, then perform focal calculation to derive
autocovariate (same procedure as for GLM)
```

```
rast_brt <- raster(ncol=39, nrow = 47, ymn = 1, ymx = 40, xmn = 1, xmx = 50)
```

```
res(rast_brt) <- 1
```

```
xy_res_brt <- cbind(xy, resid(env_brt))
```

```
rast_brt[cellFromXY(rast_brt, xy_res_brt)] <- xy_res_brt[,3]
```

```
focal_rac_rast_brt <- focal(rast_brt, ngb = 3, fun = mean, na.rm = TRUE)
```

```
plot(focal_rac_rast_brt)
```

```
focal_rac_vect_brt <- xyValues(focal_rac_rast_brt, xy)
```

```
Snouter<- cbind(Snouter, focal_rac_vect_brt)
```

```
    #Fit the BRT RAC model
```

```
    rac_brt <- gbm.step(data = Snouter, gbm.x = c(3:4,17), gbm.y = 5, family = "bernoulli",
tree.complexity = 3, learning.rate = 0.002, bag.fraction = 0.5)
```

```
#####
```

```
#Autolog model, use the autocovariate based on response variable value (already calculated for
```

```
#GLMs), previously appended to Snouter as column16
```

```
auto_brt <- gbm.step(data = Snouter, gbm.x = c(3:4,16), gbm.y = 5, family = "bernoulli",
tree.complexity = 3, learning.rate = 0.002, bag.fraction = 0.5)
```

```
#####
```