Supplementary Information for:

Do long-distance migratory birds track their niche through seasons?

Damaris Zurell, Laure Gallien, Catherine H. Graham, and Niklaus E. Zimmermann

Table of contents:

Supplementary Figures

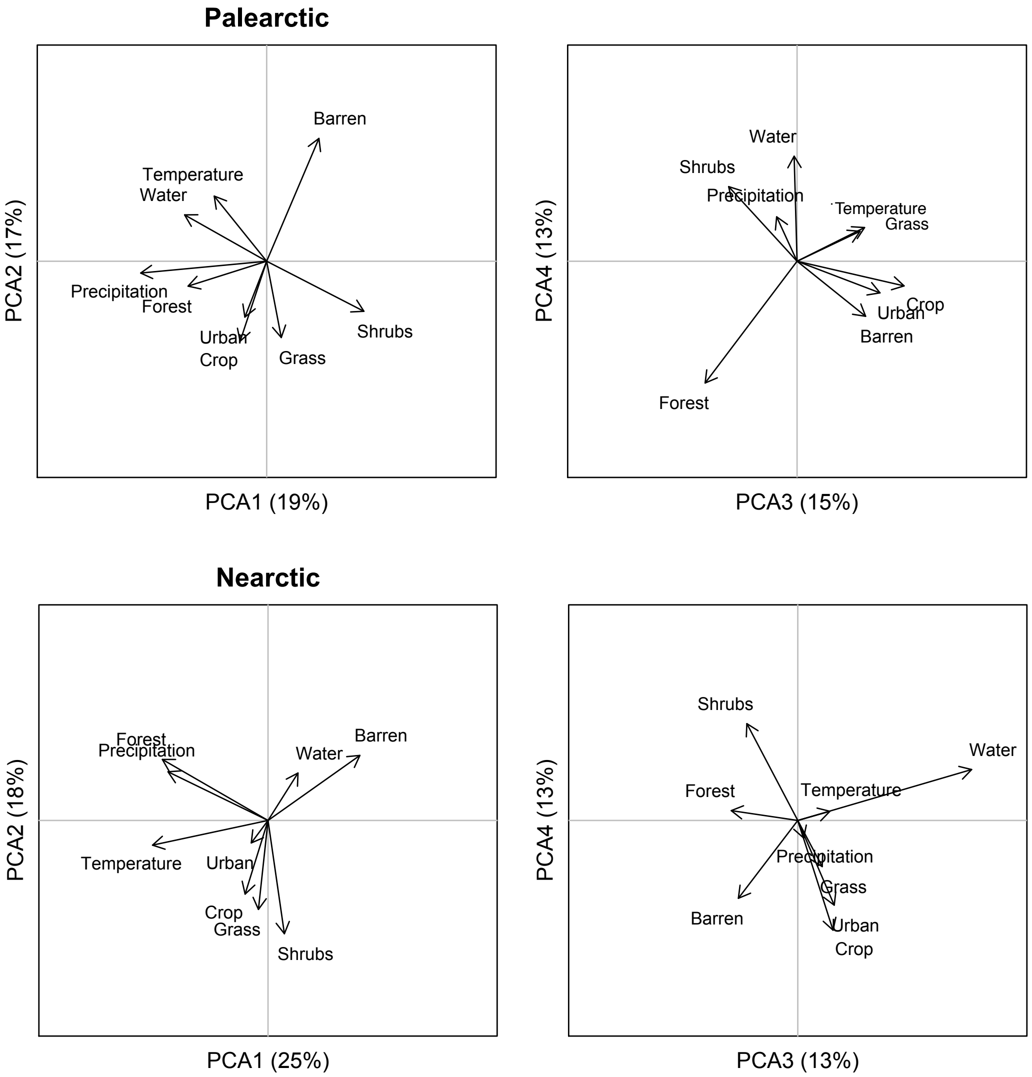

Figure S1. Principal component analyses on climate and land cover variables in Western and Eastern Hemisphere. Shown are the loadings for the first four PCA axes.

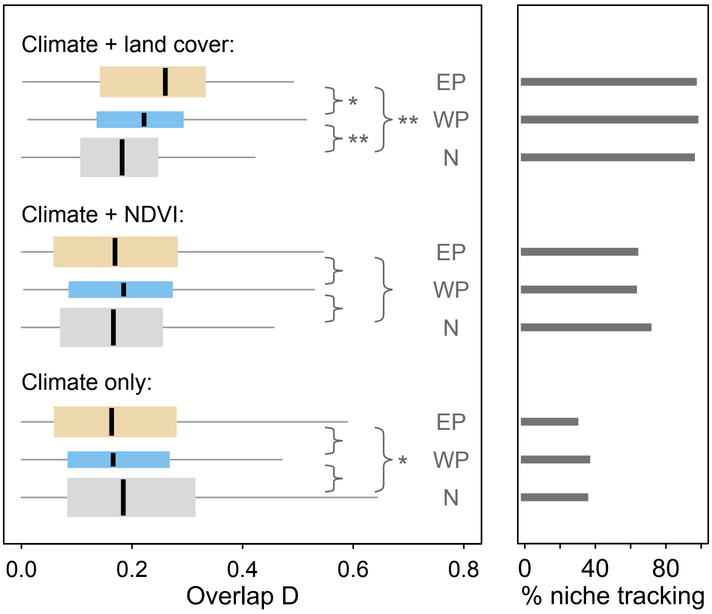

Figure S2. Niche overlap *D* between summer breeding and winter ranges (left) and proportion of migratory bird species significantly tracking their seasonal niche (right). We distinguish long-distance migratory birds breeding in Nearctic (N), Western Palearctic (WP) and Eastern Palearctic (EP). Niche overlap *D* was estimated along niche axes obtained from PCA for the three predictor sets climate only, climate + NDVI, and climate + land cover; outliers are not displayed in the boxplot. Asterisks indicate significant differences following a two-tailed Wilcoxon rank sum test (p < 0.01; * p < 0.05). The width of the boxplots is proportional to the species numbers within geographic regions (N *n*=329; WP *n*=132; EP *n*=256). The barplot (right) indicate results from the similarity tests where the niche (in PCA space) during summer is compared to random niches during winter and vice versa.**

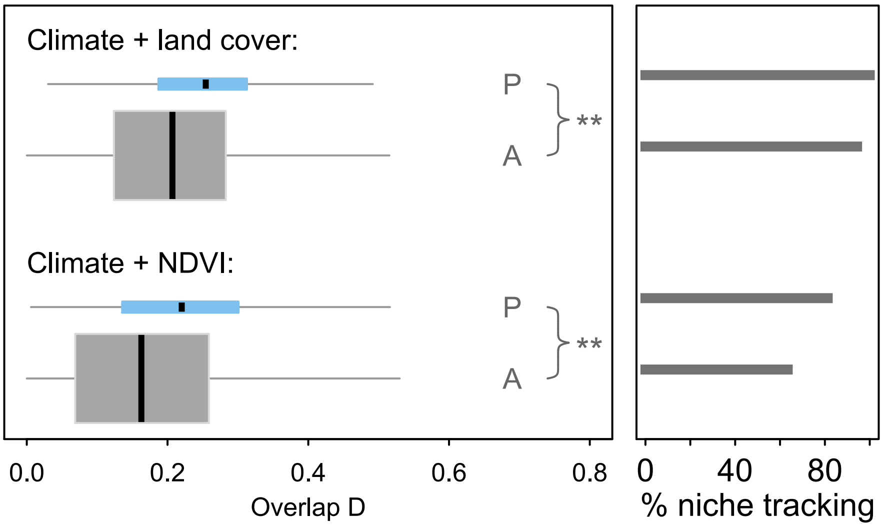

Figure S3. Niche overlap *D* (left) and proportion of significant niche tracking (right) between summer breeding and winter ranges of long-distance migratory birds employing active (A) and passive (P) flight strategies. We broadly classified all species that regularly employ soaring flight, even if not exclusively, as passive flyers. Niche overlap *D* was estimated along niche axes obtained from PCA for the two different predictor sets climate + NDVI (using all PCA axes), and climate + land cover (using the first four PCA axes). Asterisks indicate significant differences following a two-tailed Wilcoxon rank sum test (p < 0.01; * p < 0.05). The width of the boxplots is proportional to the species numbers within each flight category (n=637 for active and n=80 for passive flight strategies). The colours in the barplot (right) indicate results from the similarity tests where the niche (in PCA space) during summer is compared to random niches during winter (red) or the niche during winter is compared to random niches during summer (blue).**

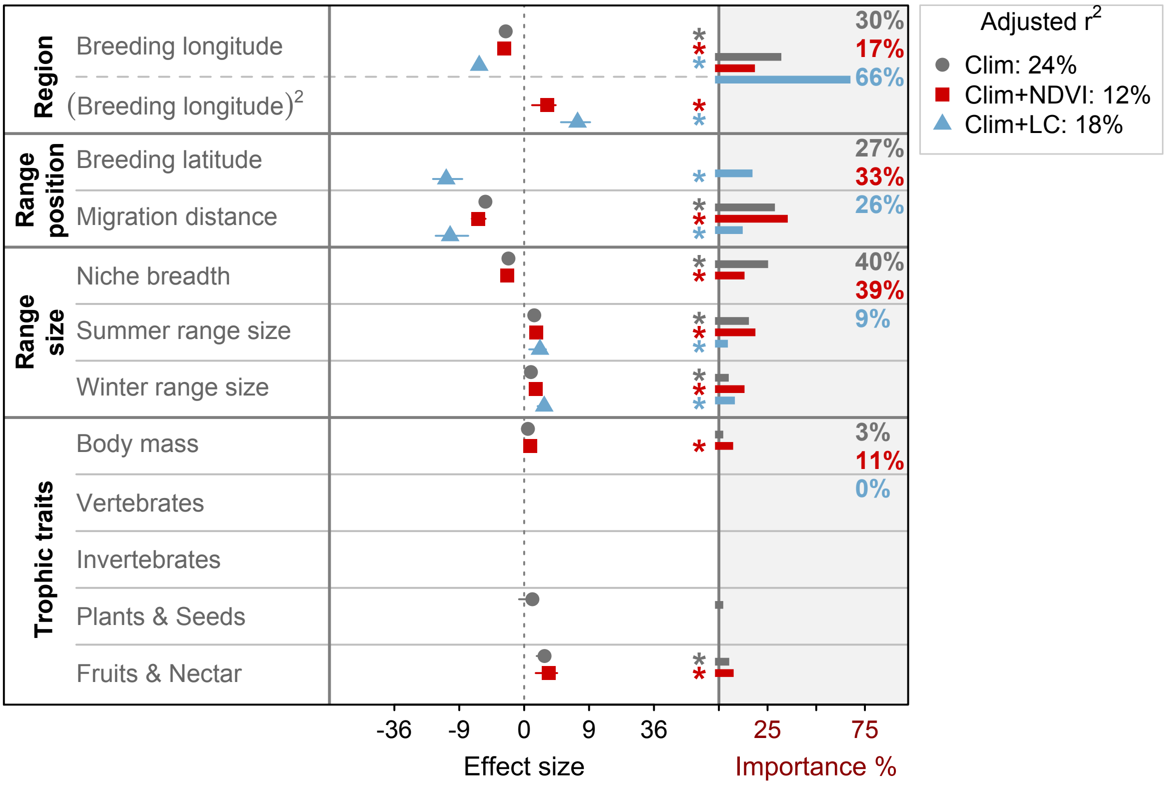

Figure 4. Relationship of migratory bird species niche tracking with biogeographic and ecological factors. Niche tracking is given by the standardised effect size *SES*$_D$ estimated from niche similarity tests. To explain *SES*$_D$ values from traits, multivariate models were estimated by phylogenetic regression using AIC-based stepwise variable selection (n=717). Asterisks indicate significant coefficients (p<0.05). Bars indicate importance of each variable selected in the final model (note that linear and quadratic terms of longitude are summarised into a single importance value), and given percentages sum the variable importance for the four different categories of biogeographic and ecological factors (cf. Table 1). Overall explained variance is given by the adjusted r^2 of the multivariate model.

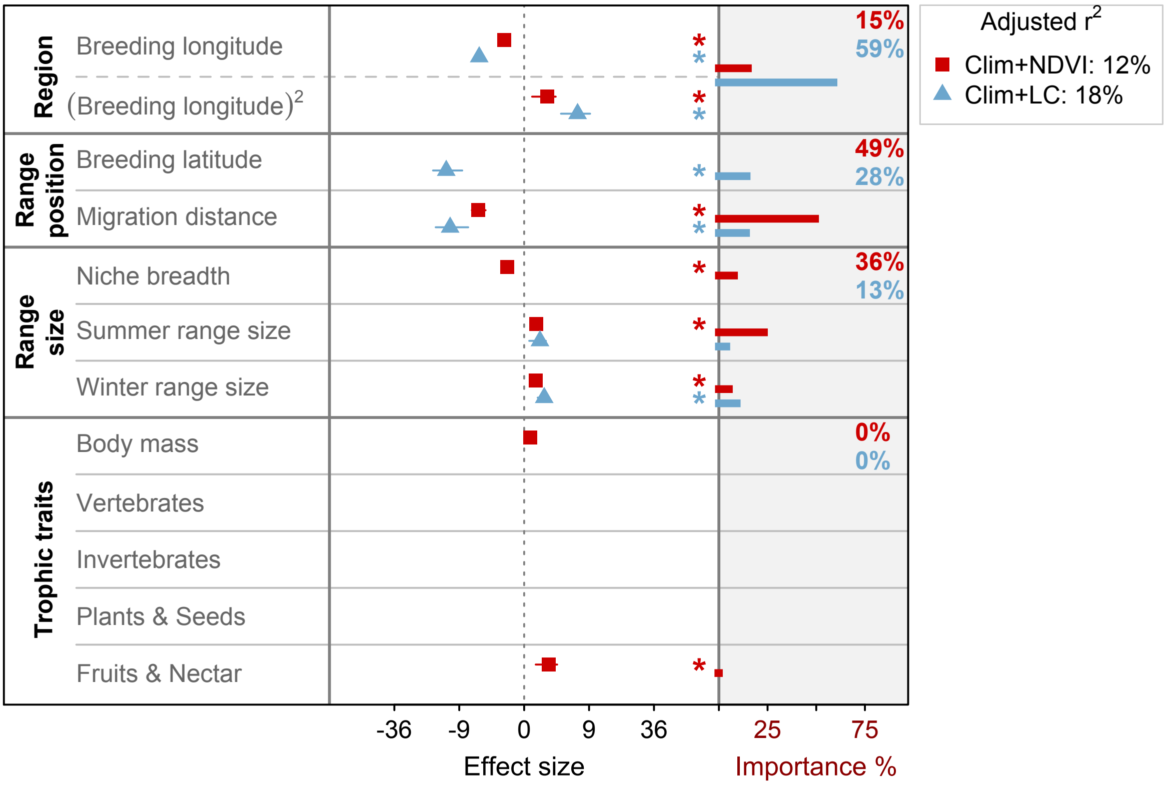

Figure S5. Relationship of migratory bird species niche tracking with geographical and ecological factors for long-distance migrants employing an active flight mode (n=637). First, niche overlap *D* was calculated using ordination (PCA) on the two different predictor sets climate + NDVI (red) and climate + land cover (blue). Then, similarity tests were conducted to test for deviation of *D* from random expectation (with 200 permutations) by comparing the niche during summer against random niches during winter and vice versa. Niche tracking was given by the standardised effect sizes defined as SES_D = (observed *D* – mean of simulated *D*)/standard deviation of simulated *D*. Multivariate models were then estimated by phylogenetic regression using AIC-based stepwise variable selection to explain SES_D values from traits. Asterisks indicate significant coefficients ($p<0.05$). Bars indicate importance of each variable in the regression (please note that linear and quadratic terms of the same variable are summarised into a single importance value), and given percentages sum the variable importance for the four different categories of geographic and ecological factors (cf. Table1 in main article). Overall explained variance is given by the adjusted r^2 of the multivariate model.

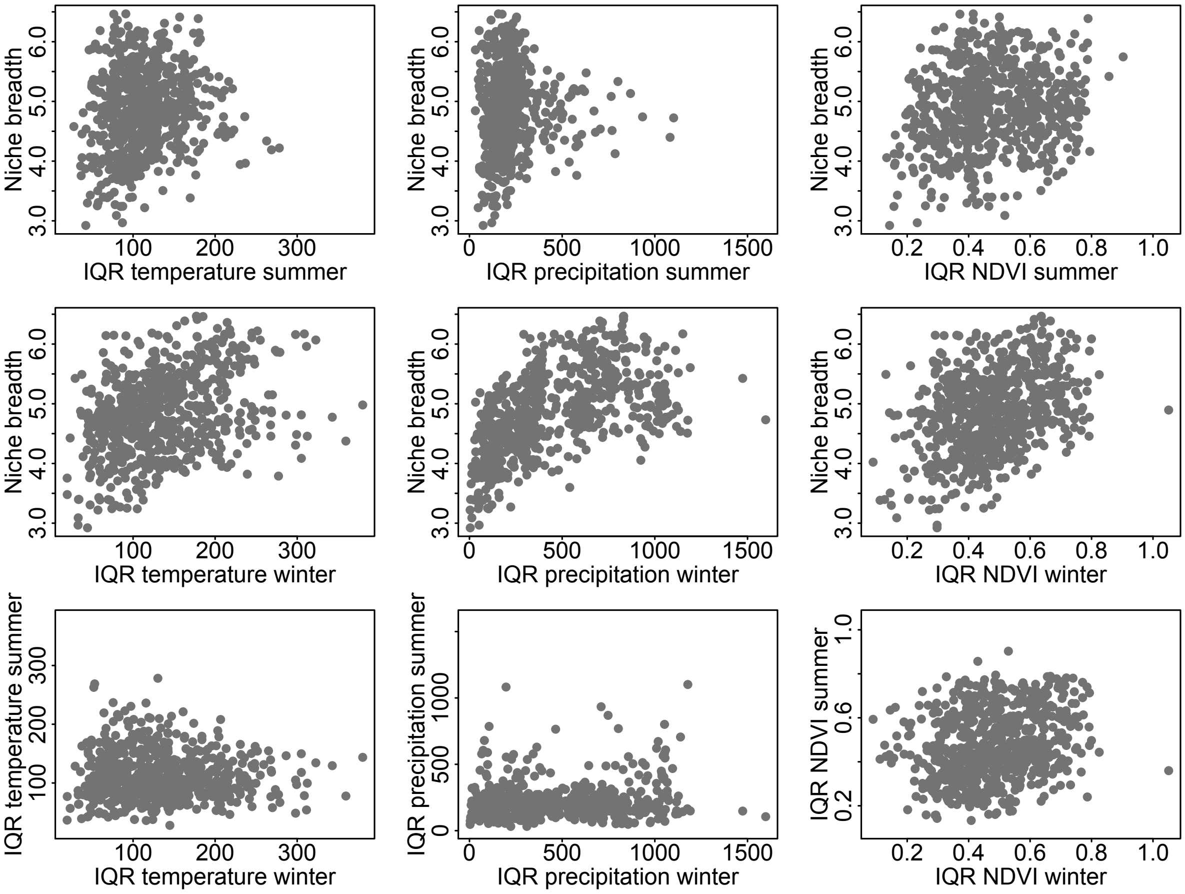

Figure S6. Relationship between total annual niche breadth of long-distance migratory bird species for climate and vegetation productivity (NDVI) and 90 % inter-quantile ranges of the single environmental variables (temperature, precipitation and NDVI) in summer and winter.

Supplementary Tables

Table S1. Land cover classification. The original UMD land cover classes were aggregated to seven broad land cover classes.

UMD land cover classification	Aggregated land cover classes
Water	Water
Evergreen needle-leaf forest	Woodlands
Evergreen broad-leaf forest	
Deciduous needle-leaf forest	
Deciduous broadleaf forest	
Mixed forest	
Woodland	
Wooded grassland	Shrublands
Closed shrubland	
Open shrubland	
Grassland	Grassland
Cropland	Cropland
Bare ground	Bare ground
Urban and built	Urban and built

R Codes for multi-dimensional niche overlap analyses

```
## Extension of the niche overlap analyses in the R package ecospat
## adapted by D. Zurell (Swiss Federal Research Institute WSL, contact: 
damaris.zurell@wsl.ch),
## to allow for multi-dimensional niche estimation using the R package 'ks'.
## implemented in February 2017, last changed November 2017
###################################################################################
###############

# I call the ecospat package first so that all dependencies and imported packages 
are loaded (although the code actually just uses the niche.overlap function from 
ecospat). I have written a new version of the function for 'ecospat.grid.clim.dyn' 
called 'grid.clim.hypervolume', and a new version of the function 
'ecospat.niche.similarity.test' called 'niche.similarity.test.hypervolume' that 
work generically for up to 4 data dimensions.

library(ecospat)
library(ks)

grid.clim.hypervol <- function (glob, glob1, sp, R, th.sp = 0, th.env = 0)
{
    glob <- as.matrix(glob)
    glob1 <- as.matrix(glob1)
    sp <- as.matrix(sp)
    l <- list()

    if (ncol(glob) > 4)
        stop("Calculate overlap with more than four axes is not tested yet; the 
binning procedure won't work")

       require(ks)

        xmin <- apply(glob,2,min)
        xmax <- apply(glob,2,max)

        gr = R

        grid <- 
sapply(seq_len(ncol(glob)),FUN=function(x){seq(xmin[x],xmax[x],length=gr)})

        glob1r =glob1; spr = sp

        sp.dens <- 
kde(spr,binned=T,bgridsize=rep(gr,ncol(spr)),xmin=xmin,xmax=xmax)$estimate

        z <- sp.dens*nrow(sp)/sum(sp.dens)          #rescale density to the number 
of occurrences in sp

        glob1.dens <- 
kde(glob1r,binned=T,bgridsize=rep(gr,ncol(glob1r)),xmin=xmin,xmax=xmax)$estimate
        Z <- glob1.dens*nrow(glob1)/sum(glob1.dens)        #rescale density to the 
number of sites in glob1

        Z[Z<max(Z)/1000] <- 0
        z[z<max(z)/1000] <- 0    # remove infinitesimally small number generated by 
kernel density function
```

```
        z[Z==0] <- 0

        z.uncor <- z/max(z)      # rescale between [0:1] for comparison with other
species
        w <- z.uncor
        w[w > 0] <- 1
        z.cor <- z/Z
        z.cor[is.na(z.cor)] <- 0
        z.cor <- z.cor/max(z.cor)
        l$grid <- grid
        l$z <- z
        l$z.uncor <- z.uncor
        l$z.cor <- z.cor
        l$Z <- Z
        l$glob <- glob
        l$glob1 <- glob1
        l$sp <- sp
        l$w <- w

    return(l)
}


#-----------------------------------------------
niche.similarity.test.hypervol <- function (z1, z2, reps, alternative = "greater",
rand.type=1) {
    R <- nrow(z1$grid)
    l <- list()
     if (ncol(z1$grid)>1)  {
      dim.z <- dim(z1$Z)
      shift.center <- function(center,rand.center)
{t(apply(rand.center,1,FUN=function(x){x-center}))}
      } else {
             dim.z <- dim(z1$grid)
             coords <- function(x,dim) {x}
             shift.center <- function(center,rand.center) {shift = rand.center[,1]-
center[,1]; as.matrix(shift,dim.z)}
             }
    grid.i <-
expand.grid(data.frame(matrix(seq_len(nrow(z1$grid)),nrow=nrow(z1$grid),ncol=ncol(z
1$grid))))

     obs.o<-ecospat.niche.overlap(z1,z2,cor=T)
             #observed niche overlap
     sim.o<-data.frame(matrix(nrow=reps,ncol=2))
     #empty list of random niche overlap
     names(sim.o)<-c("D","I")

     center <- arrayInd(which(z2$z.cor==1),dim.z)                        # define
the centroid of the observed niche
     rand.center <-
arrayInd(sample(seq_len(prod(dim.z)),size=reps,prob=z1$Z/max(z1$Z),replace=F),dim.z
)            # randomly (weighted by environment prevalence) define the new
centroid for the niche
     xshift <- shift.center(center,rand.center)

     sim.o1 <- sapply(seq_len(reps),FUN=function(i){
      grid.shift <- grid.i-xshift[i,]
      i.shift <- rowSums(grid.shift < 1 | grid.shift > nrow(z1$grid)) < 1
      z2.sim <- z2
      z2.sim$z.cor[] <- 0
      z2.sim$z.cor[as.matrix(grid.shift[i.shift,])] <-
z2$z.cor[as.matrix(grid.i[i.shift,])]
      z2.sim$z.cor <- (z2$Z!=0)*1*z2.sim$z.cor                              #
remove densities out of existing environments

      if (sum(z2.sim$z.cor)) ecospat.niche.overlap(z1,z2.sim,cor=T) else
list(D=0,I=0)
     })
     sim.o <- data.frame(D=unlist(sim.o1['D',]),I=unlist(sim.o1['I',]))

    if (rand.type==2) {
        center <- arrayInd(which(z1$z.cor==1),dim.z)                    # define
the centroid of the observed niche
        rand.center <-
arrayInd(sample(seq_len(prod(dim.z)),size=reps,prob=z2$Z/max(z2$Z),replace=F),dim.z
)            # randomly (weighted by environment prevalence) define the new
centroid for the niche
        xshift <- shift.center(center,rand.center)

        sim.o2 <- sapply(seq_len(reps),FUN=function(i){
            grid.shift <- grid.i-xshift[i,]
            i.shift <- rowSums(grid.shift < 1 | grid.shift > nrow(z2$grid)) < 1
            z1.sim <- z1
            z1.sim$z.cor[] <- 0
            z1.sim$z.cor[as.matrix(grid.shift[i.shift,])] <-
z1$z.cor[as.matrix(grid.i[i.shift,])]
            z1.sim$z.cor <- (z1$Z!=0)*1*z1.sim$z.cor
     # remove densities out of existing environments

            if (sum(z1.sim$z.cor)) ecospat.niche.overlap(z2,z1.sim,cor=T) else
list(D=0,I=0)
        })
        sim.o <-
rbind(sim.o,data.frame(D=unlist(sim.o2['D',]),I=unlist(sim.o2['I',])))

    }


     l$sim<-sim.o
     # storage
     l$obs<-obs.o
     # storage

    if (alternative == "greater") {
        l$p.D <- (sum(sim.o$D >= obs.o$D) + 1)/(length(sim.o$D) +
            1)
        l$p.I <- (sum(sim.o$I >= obs.o$I) + 1)/(length(sim.o$I) +
            1)
    }
    if (alternative == "lower") {
        l$p.D <- (sum(sim.o$D <= obs.o$D) + 1)/(length(sim.o$D) +
            1)
        l$p.I <- (sum(sim.o$I <= obs.o$I) + 1)/(length(sim.o$I) +
            1)
    }

     l$ses.rank.D <- ((sum(obs.o$D > sim.o$D)+1)/(length(sim.o$D)+1))   #
standardised effect size based on the rank of the observation among simulated
values
     l$ses.rank.I <- ((sum(obs.o$I > sim.o$I)+1)/(length(sim.o$I)+1))
     l$ses.z.D <- (obs.o$D - mean(sim.o$D)) / sd(sim.o$D)                      #
standardised effect size calculated as standardised mean difference between
observation and simulated values (the z-score)
     l$ses.z.I <- (obs.o$I - mean(sim.o$I)) / sd(sim.o$I)

     return(l)
}
```