ecoevo-hmsc

```
library(ecoevor)
library(here)
library(R.matlab)
library(grDevices)
library(knitr)
library(Hmsc)
library(beanplot)
library(dplyr)
library(gplots)
```

1 Overview

The goal of this vignette is to illustrate how to use HMSC (Heirarchical Models of Species Communities; Ovaskainen et al. 2017; Tikhonov et al. 2020) to analyze eco-evolutionary community and metacommunity data. The eco-evolutionary data is generated using a multi-species model of growth and competition (a Leslie-Gower model, which is a discrete-time Lotka-Volterra model). In this model, we also include the possibility that the population growth rate can evolve (using a model of evolutionary rescue introduced by Gomulkiewicz & Holt 1995). The model is as follows:

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1 + \alpha_{ii} N_{i,t} + \sum_j \alpha_{ij} N_{j,t}}$$

where N_i, t is the population size of a species at time t,

$$\hat{W}$$

is calculated as:

$$\hat{W} = W_{max} \sqrt{(\frac{w}{P+w})}$$

,

W_{max} is the species' maximum per-capita growth rate, w is the width of the Gaussian fitness function (which determines the strength of selection), P is is the width of the distribution of the phenotype x, h^2 is the heritability of the trait x, E is the local environmental optimum trait value, x_t is the trait value of the species at time t, α_{ii} is the intraspecific competition coefficient and α_{ij} is the interspecific competition coefficient.

I consider the evolution of multiple species ($j = 15$) in a landscape of patches ($i = 50$). The patches have values of an environmental property E (drawn from a uniform distribution ranging from 0 to 1) that determines the local optimum phenotype E, i.e. where species experience their absolute fitness W_{max}, and patches also have spatial locations X and Y (both drawn from U(0, 1)). The patches thus have a connectivity matrix $\mathbf{D}$ (here given by their Euclidean distance), as well as a connectivity matrix $\mathbf{C}$ that is a Gaussian function of $\mathbf{D}$ and a species dispersal rate d_j.

2 Simulation

The simulation is run in MATLAB (the code for this can be found at https://github.com/jhpantel/HMSC_ecoevo/code/simulation/metacom_evol_sim_same_init.m), and generates a 50x15x1000 array of abundance values (for 50 patches, 15 species, at 1000 time points). In this example, we consider $h^2 = 0.1$ and $d = 10^{-3}$ for all 15 species.

We can see the dynamics of the simulation here:

```
datapath <- here::here("inst","extdata", "hmsc")
result <- "h_01_d_minus3"  # Enter desired simulation conditions here
res <- R.matlab::readMat(paste(datapath,"/",result,"_res.mat",sep=""))
N <- res$N

## Plot of Patch 1 community dynamics over time
cl <- grDevices::colors(distinct = TRUE)
mycols2 <- sample(cl, 15)

par(mfrow=(c(10,5)))
par(mar=c(1,1,1,1))
for(j in 1:50){
  plot(res$N[j,1,],col=mycols2[1],ylim=c(0,800),axes=FALSE,type="n",ylab=NA,xlab=NA,xaxs="i",yaxs="i")
  if (j %in% c(1,6,11,16,21,26,31,36,41,46))
    axis(2,col="grey40",col.axis="grey20",at = c(0,250,500,750,1000))
  if (j %in% c(46,47,48,49,50))
    axis(1,col="grey40",col.axis="grey20",at = c(0,200,400,600,800))
  for(i in 1:15){
    points(res$N[j,i,],col=mycols2[i])
  }
}
```

The plots are time series of population size (y-axis) over time (x-axis) for all species in all 50 patches. The plots are ordered by rows, from lowest to highest E values and each species is plotted with a unique color. Traits are evolving for all species as well, which we can visualize here:

```
par(mfrow=(c(10,5)))
par(mar=c(1,1,1,1))
for(j in 1:50){
  plot(res$xt[j,1,],col=mycols2[1],ylim=c(-2,1),axes=FALSE,type="n",ylab=NA,xlab=NA,xaxs="i",yaxs="i")
  if (j %in% c(1,6,11,16,21,26,31,36,41,46))
    axis(2,col="grey40",col.axis="grey20",at = c(-2,-1.5,-1,-.5,0,.5,1))
  if (j %in% c(46,47,48,49,50))
    axis(1,col="grey40",col.axis="grey20",at = c(0,250,500,750,1000))
  for(i in 1:15){
    points(res$xt[j,i,],col=mycols2[i])
```

```
  }
}
```

3 HMSC Analysis

The goal of our analysis is to estimate the relative effects of intrinsic dopulationn dynamics, environmental drivers, spatial structure, and trait evolution for driving community composition in the metacommunity. We use HMSC to build a linear Bayesian hierarchical model.

To analyze our data with HMSC, we first re-arrange our data to fit the structure needed for input into HMSC.

```r
# Arrange all data for HMSC
## site-by-species abundance data as a time series
N <- res$N
Y <- array(NA,dim=c(50000,15),dimnames=list(NULL,c("y1","y2","y3","y4","y5","y6","y7","y8","y9","y10","y
count <- seq(1,50000,by=50)
for(i in 1:1000){
  Y[(count[i]:(count[i]+49)),] <- N[,,i]
}

## site-by-species trait data as a time series
trait_x <- res$xt
trait_init <- trait_x[,,1]  # Eliminate my initial trait value of 9999 for absent species
trait_init[trait_init == 9999] <- NaN
trait_x[,,1] <- trait_init
## site-by-species CHANGE in trait value as a time series
delta_x <- array(NA,c(50,15,999),dimnames=list(NULL,c("x1","x2","x3","x4","x5","x6","x7","x8","x9","x10
for(i in 2:1000){
  delta_x[,,i-1] <- trait_x[,,i] - trait_x[,,i-1]
}
delta_x_time <- array(NA,dim=c(49950,15),dimnames=list(NULL,c("x1","x2","x3","x4","x5","x6","x7","x8","x
count <- seq(1,49950,by=50)
for(i in 1:999){
  delta_x_time[(count[i]:(count[i]+49)),] <- delta_x[,,i]
}
delta_x_time[is.nan(delta_x_time)] <- 0

## Random effects of site and year
Random <- array(NA,dim=c(50000,2),dimnames=list(NULL,c("site","year")))
Random[,1] <- rep(1:50,1000)
Random[,2] <- rep(1:1000,each=50)
Random <- as.data.frame(Random)

## Fixed effects: environment, site abundance the year before, species trait value the year before
X <- Y[Random$year != 1000,]  # Use abundance in years 1-999 as fixed effects - the abundance the year
Y <- Y[Random$year != 1,]  # Trim Y to only include years 2-1000

Random <- Random[Random$year != 1,]  # Trim Random to only include years 2-1000
Random$site <- as.factor(Random$site)
Random$year <- as.factor(Random$year)
Random$sample <- as.factor(1:nrow(Random))


## Random effects
xy <- res$xy
rownames(xy) <- 1:50
rL.spatial <- Hmsc::HmscRandomLevel(sData = xy)
rL.spatial = Hmsc::setPriors(rL.spatial,nfMin=1,nfMax=5)

rL.time <- Hmsc::HmscRandomLevel(units = unique(Random$year))
rL.time = Hmsc::setPriors(rL.time,nfMin=1,nfMax=5)

## Fixed effects: environment
env_hold <- array(NA,dim=c(dim(Y)[1],dim(res$E)[2]))
```

```r
for(i in 1:dim(res$E)[2]){ # for each environmental variable
  env_hold[,i] <- rep(res$E[,i],999)
  colnames(env_hold)[i] <- paste("env",i,sep="")
}

## Interaction effects: site abundance the year before X change in species trait value
X_delta_x <- X * delta_x_time
colnames(X_delta_x) <- paste("x",colnames(X_delta_x),sep="")

# When h2 = 0
#X <- cbind(X,env_hold)
#X <- scale(X)
#X <- cbind(X,MEM_hold)

# When h2 != 0
X <- cbind(X,env_hold,delta_x_time,X_delta_x)
X <- scale(X)
#X <- cbind(X,MEM_hold)
X <- as.data.frame(X)

m <- Hmsc::Hmsc(Y=Y,XData=X,studyDesign=Random,ranLevels=list("site"=rL.spatial,"year"=rL.time),distr="
# The model takes a long time to run. The results included here are not reliable, but demonstrate how t
m.1 <- Hmsc::sampleMcmc(m,10,transient=5,nChains=2,nParallel=2,verbose=5)
```

Now that the model is fit, we can analyze the results.

```r
#Explanatory power
preds = Hmsc::computePredictedValues(m.1)
MF = Hmsc::evaluateModelFit(hM=m.1, predY=preds)
```

We should also evaluate the chains, for convergence. Then we will inspect the results. The resulting coda object can be inspected as in the Hmsc vignette, using coda posterior commands such as plot.mcmc and gelman.diag. However, we have 705 predictors, and these will be too large to view easily. We instead explain the structure of the resulting Hmsc object.

α_{ij}

For each of the 15 species, we have the following regression coefficient estimates: * 1. intercept * 2-16. the impacts of $N_{j,t-1}$ for $N_{i,t}$, for j = 1-15) + e.g. for species 1: 2 is the impact of $N_{1,t-1}$ for $N_{1,t}$ + 3. the impact of $N_{2,t-1}$ for $N_{1,t}$ + 4. the impact of $N_{3,t-1}$ for $N_{1,t}$ + 5. the impact of $N_{4,t-1}$ for $N_{1,t}$ + 6. the impact of $N_{5,t-1}$ for $N_{1,t}$ + 7. the impact of $N_{6,t-1}$ for $N_{1,t}$ + 8. the impact of $N_{7,t-1}$ for $N_{1,t}$ + 9. the impact of $N_{8,t-1}$ for $N_{1,t}$ + 10. the impact of $N_{9,t-1}$ for $N_{1,t}$ + 11. the impact of $N_{10,t-1}$ for $N_{1,t}$ + 12. the impact of $N_{11,t-1}$ for $N_{1,t}$ + 13. the impact of $N_{12,t-1}$ for $N_{1,t}$ + 14. the impact of $N_{13,t-1}$ for $N_{1,t}$ + 15. the impact of $N_{14,t-1}$ for $N_{1,t}$ + 16. the impact of $N_{15,t-1}$ for $N_{1,t}$ * 17. The impact of environment (E) for $N_{i,t}$ * 18-32. the impacts of $x_{j,t-1}$ for $N_{i,t}$, for j = 1-15) + e.g. for species 1: 18 is the impact of $x_{1,t-1}$ for $N_{1,t}$ + 19. the impact of $x_{2,t-1}$ for $N_{1,t}$ + 20. the impact of $x_{3,t-1}$ for $N_{1,t}$ + 21. the impact of $x_{4,t-1}$ for $N_{1,t}$ + 22. the impact of $x_{5,t-1}$ for $N_{1,t}$ + 23. the impact of $x_{6,t-1}$ for $N_{1,t}$ + 24. the impact of $x_{7,t-1}$ for $N_{1,t}$ + 25. the impact of $x_{8,t-1}$ for $N_{1,t}$ + 26. the impact of $x_{9,t-1}$ for $N_{1,t}$ + 27. the impact of $x_{10,t-1}$ for $N_{1,t}$ + 28. the impact of $x_{11,t-1}$ for $N_{1,t}$ + 29. the impact of $x_{12,t-1}$ for $N_{1,t}$ + 30. the impact of $x_{13,t-1}$ for $N_{1,t}$ + 31. the impact of $x_{14,t-1}$ for $N_{1,t}$ + 32. the impact of $x_{15,t-1}$ for $N_{1,t}$ * 33-47. the impacts of $x_{j,t-1} \cdot N_{j,t-1}$ for $N_{i,t}$, for j = 1-15) $x_{1,t-1} \cdot N_{1,t-1}$ for $N_{1,t}$ + e.g. for species 1: 33 is the impact of $x_{1,t-1} \cdot N_{1,t-1}$ for $N_{1,t}$ + 34. the impact of $x_{2,t-1} \cdot N_{2,t-1}$ for $N_{1,t}$ + 35. the impact of $x_{3,t-1} \cdot N_{3,t-1}$ for $N_{1,t}$ + 36. the impact of $x_{4,t-1} \cdot N_{4,t-1}$ for $N_{1,t}$ + 37. the impact of $x_{5,t-1} \cdot N_{5,t-1}$ for $N_{1,t}$ + 38. the impact of $x_{6,t-1} \cdot N_{6,t-1}$ for $N_{1,t}$ + 39. the impact of $x_{7,t-1} \cdot N_{7,t-1}$ for $N_{1,t}$ + 40. the impact of $x_{8,t-1} \cdot N_{8,t-1}$ for $N_{1,t}$ + 41. the impact of

$x_{9,t-1} \cdot N_{9,t-1}$ for $N_{1,t} + 42$. the impact of $x_{10,t-1} \cdot N_{10,t-1}$ for $N_{1,t} + 43$. the impact of $x_{11,t-1} \cdot N_{11,t-1}$ for $N_{1,t} + 44$. the impact of $x_{12,t-1} \cdot N_{12,t-1}$ for $N_{1,t} + 45$. the impact of $x_{13,t-1} \cdot N_{13,t-1}$ for $N_{1,t} + 46$. the impact of $x_{14,t-1} \cdot N_{14,t-1}$ for $N_{1,t} + 47$. the impact of $x_{15,t-1} \cdot N_{15,t-1}$ for $N_{1,t}$

This repeats for all $j = 15$ species, i.e. coefficients 48-94 are for species 2, coefficients 95-141 are for species 3, etc.

```
m.post = Hmsc::convertToCodaObject(m.1)
m.df <- as.data.frame(rbind(m.post$Beta[[1]],m.post$Beta[[2]]))
## intercept
B.I <- m.df[,seq(1,705,by=47)]
beanplot::beanplot(B.I,log="")
```

```
## Fixed effects: environment
B.E <- m.df[,seq(17,705,by=47)]
beanplot::beanplot(B.E)
```

```
## Fixed effects - Species ABUNDANCE T-1 association plots - Influence of Yt-1 for Yt
B.N <- m.df[,2:16]
for(i in 1:14){
  B.N <- cbind(B.N,m.df[,(2+(i*47)):(16+(i*47))])
}
# Mean and 95% highest density interval for regression coefficients
B.N.mean <- apply(B.N,2,mean)
B.N.hdi <- apply(B.N,2,quantile,probs=c(0.025,.975))
support <- apply(B.N.hdi,2,function(x) dplyr::between(0,x[1],x[2]))
# Plot posterior mean if HDI does not include 0
B.N.mean.support <- B.N.mean
B.N.mean.support[support == T] <- 0

data <- matrix(B.N.mean.support,15,15)
#diag(data) <- 0
rnames <- colnames(X[1:15])
my_palette <- colorRampPalette(c("blue", "white", "red"))(200)
col_breaks = seq(-3,3,length.out=201)
xl <- expression(N["i,t"] ~ "(response, y)")
yl <- expression(beta ~ "*" ~ N["j,t-1"] ~ "(driver, x)")
t <- expression("Posterior mean," ~ {0 %notin% HDI})
diag_val <- round(data)
diag_val[upper.tri(diag_val)] <- NA
diag_val[lower.tri(diag_val)] <- NA
```

```r
heatmap.2(data,
          Rowv=FALSE,
          cellnote = diag_val,  # same data set for cell labels
          main = t ,# heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(5,5),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          breaks=col_breaks,    # enable color transition at specified limits
          dendrogram="none",     # only draw a row dendrogram
          xlab=xl,
          ylab=yl,
          Colv="NA")            # turn off column clustering
```

```r
## Fixed effects - Species TRAIT association plots - Influence of X at t-1 on abundance at time t
B.dx <- m.df[,18:32]
for(i in 1:14){
  B.dx <- cbind(B.dx,m.df[,(18+(i*47)):(32+(i*47))])
}
# Mean and 95% highest density interval for regression coefficients
B.dx.mean <- apply(B.dx,2,mean)
B.dx.hdi <- apply(B.dx,2,quantile,probs=c(0.025,.975))
support <- apply(B.dx.hdi,2,function(x) dplyr::between(0,x[1],x[2]))
# Plot posterior mean if HDI does not include 0
```

```r
B.dx.mean.support <- B.dx.mean
B.dx.mean.support[support == T] <- 0

data <- matrix(B.dx.mean.support,15,15)
#diag(data) <- 0
rnames <- colnames(X[1:15])
my_palette <- colorRampPalette(c("blue", "white", "red"))(200)
col_breaks = seq(-4,4,length.out=201)
xl <- expression(N["i,t"] ~ "(response, y)")
yl <- expression(beta ~ "*" ~ N["j,t-1"] ~ "(driver, x)")
t <- expression("Posterior mean," ~ {0 %notin% HDI})
diag_val <- round(data)
diag_val[upper.tri(diag_val)] <- NA
diag_val[lower.tri(diag_val)] <- NA

heatmap.2(data,
          Rowv=FALSE,
          #cellnote = data,  # same data set for cell labels
          main = t ,# heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(5,5),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          breaks=col_breaks,    # enable color transition at specified limits
          dendrogram="none",     # only draw a row dendrogram
          xlab=xl,
          ylab=yl,
          Colv="NA")            # turn off column clustering
```

```
## Fixed effects - Species N_deltaX interaction association plots - Influence of X * delta_x at t-1 on
B.N_dx <- m.df[,33:47]
for(i in 1:14){
  B.N_dx <- cbind(B.N_dx,m.df[,(33+(i*47)):(47+(i*47))])
}
# Mean and 95% highest density interval for regression coefficients
B.N_dx.mean <- apply(B.N_dx,2,mean)
B.N_dx.hdi <- apply(B.N_dx,2,quantile,probs=c(0.025,.975))
support <- apply(B.N_dx.hdi,2,function(x) dplyr::between(0,x[1],x[2]))
# Plot posterior mean if HDI does not include 0
B.N_dx.mean.support <- B.N_dx.mean
B.N_dx.mean.support[support == T] <- 0

data <- matrix(B.N_dx.mean.support,15,15)
#diag(data) <- 0
rnames <- colnames(X[1:15])
my_palette <- colorRampPalette(c("blue", "white", "red"))(200)
col_breaks = seq(-2,2,length.out=201)
xl <- expression(N["i,t"] ~ "(response, y)")
yl <- expression(beta ~ "*" ~ N["j,t-1"] ~ "(driver, x)")
t <- expression("Posterior mean," ~ {0 %notin% HDI})
diag_val <- round(data)
diag_val[upper.tri(diag_val)] <- NA
diag_val[lower.tri(diag_val)] <- NA
```

```
heatmap.2(data,
          Rowv=FALSE,
          #cellnote = data,  # same data set for cell labels
          main = t ,# heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          margins =c(5,5),     # widens margins around plot
          col=my_palette,       # use on color palette defined earlier
          breaks=col_breaks,    # enable color transition at specified limits
          dendrogram="none",     # only draw a row dendrogram
          xlab=xl,
          ylab=yl,
          Colv="NA")            # turn off column clustering
```

References

Gomulkiewicz, R. & Holt, R. D. When does evolution by natural selection prevent extinction? Evolution 49, 201–207 (1995)

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. and Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters, 20(5), 561-576.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M., Oksanen, J., & Ovaskainen, O.

(2020). Joint species distribution modelling with the R-package Hmsc. Methods in Ecology and Evolution, 11(3), 442-447.