HMSC 1st model runs

The purpose here is to run a first series of models for Crozet and explore the results I extracted using the HMSC pipeline. Here I show the results for a model using 4 chains, 250 samples, and a thinning of 1000.

## Model inputs

# Y = community pres/abs matrix of 1956 sites \* 19 species (all but Limosella australis that was too rare)  
# XData = the environmental matrix :  
XData <- models$`presence-absence model`$XData  
head(XData)

## accum\_prec mean\_temp numero\_observation jour mois annee pente exposition id  
## 1 2666.0 4.856885 C10A01 13 12 2010 24 E 1  
## 2 2666.0 4.856885 C10A02 13 12 2010 24 E 2  
## 3 2798.9 4.882104 C10A05 13 12 2010 22 E 3  
## 4 2798.9 4.810034 C10A07 13 12 2010 24 E 4  
## 5 2917.1 4.794531 C10A09 13 12 2010 23 E 5  
## 6 2917.1 4.775049 C10A10 13 12 2010 22 E 6

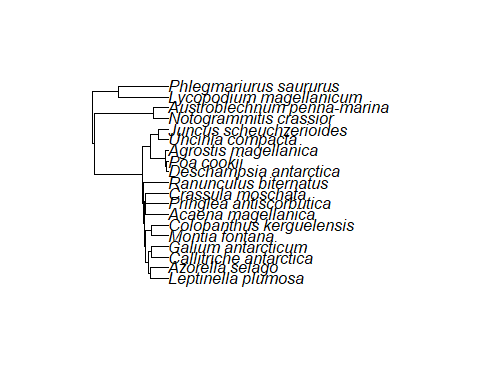
accum\_prec = Accumulated precipitation amount over 1 year (bio12 CHELSA), 0.008333 resolution (~1 km)

mean\_temp = mean temperature downscaled to ~30m, based on mean annual daily mean air temperatures averaged over 1 year (bio1 CHELSA), from 1981-2010

pente et expostion = terrain variables that come from field measures in the TAAF

jour mois annee = field sampling dates for pres/abs data

# phylogenetic tree for Crozet. I kept the phylo.maker one.  
phylotree <- ape::read.tree("../data/traits\_trees/phylomaker\_tree")  
phylotree$tip.label <- gsub("\_", " ", phylotree$tip.label)  
  
# need to remove Limosella australis for Crozet  
phylo\_cro <- ape::drop.tip(phylotree, tip = "Limosella australis")  
plot(phylo\_cro)



Still no traits, and there wil be none from TRY ==> not enough traits measured across all plant species.

HMSC also needs the followong :

# STUDY DESIGN  
studyDesign = data.frame(site=XData$numero\_observation, id=XData$id)  
  
# RANDOM EFFECT STRUCTURE, HERE Site (hierarchical study design)  
rL.site = Hmsc::HmscRandomLevel(units = levels(studyDesign$site))  
str(rL.site)

## List of 18  
## $ pi : Factor w/ 1956 levels "C10A01","C10A02",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ s : NULL  
## $ sDim : num 0  
## $ spatialMethod: NULL  
## $ x : NULL  
## $ xDim : num 0  
## $ N : int 1956  
## $ distMat : NULL  
## $ nfMax : num Inf  
## $ nfMin : num 2  
## $ nNeighbours : NULL  
## $ nu : num 3  
## $ a1 : num 50  
## $ b1 : num 1  
## $ a2 : num 50  
## $ b2 : num 1  
## $ alphapw : NULL  
## $ call : language Hmsc::HmscRandomLevel(units = levels(studyDesign$site))  
## - attr(\*, "class")= chr "HmscRandomLevel"

Not too sure what this does in detail, but it has default parameters to estimate the random effects. I didn’t alter this.

# and optionally id, if we are interested in species associations at that level  
rL.id = Hmsc::HmscRandomLevel(units = levels(studyDesign$id))  
str(rL.site)

## List of 18  
## $ pi : Factor w/ 1956 levels "C10A01","C10A02",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ s : NULL  
## $ sDim : num 0  
## $ spatialMethod: NULL  
## $ x : NULL  
## $ xDim : num 0  
## $ N : int 1956  
## $ distMat : NULL  
## $ nfMax : num Inf  
## $ nfMin : num 2  
## $ nNeighbours : NULL  
## $ nu : num 3  
## $ a1 : num 50  
## $ b1 : num 1  
## $ a2 : num 50  
## $ b2 : num 1  
## $ alphapw : NULL  
## $ call : language Hmsc::HmscRandomLevel(units = levels(studyDesign$site))  
## - attr(\*, "class")= chr "HmscRandomLevel"

Same structure as the site random level, except this one is supposed to be much finer, at the observation scale.

## Model structure

# REGRESSION MODEL FOR ENVIRONMENTAL COVARIATES.  
XFormula = ~ mean\_temp + accum\_prec + pente + exposition  
  
# REGRESSION MODEL FOR TRAITS : none.  
  
# PRESENCE-ABSENCE MODEL FOR INDIVIDUAL SPECIES (COMMON ONLY)  
m = Hmsc::Hmsc(Y=Y, XData = XData, XFormula = XFormula,  
 # TrData = TrData, TrFormula = TrFormula,  
 phyloTree = phylo\_cro,  
 distr="probit",  
 studyDesign = studyDesign, ranLevels=list(site=rL.site, id=rL.id))

## Model fit

input parameters :

* nChains = 4
* nParallel = nChains
* samples=250, thin=1000

m = Hmsc::sampleMcmc(m, samples = samples, thin=thin,  
 adaptNf=rep(ceiling(0.4\*samples\*thin),m$nr),   
 transient = ceiling(0.5\*samples\*thin),  
 nChains = nChains,  
 nParallel = nParallel)

## Results

See pdf called “results/parameter\_estimates\_ex2.pdf”).

* Variance partitioning plot: The corresponding values are in “results/parameter\_estimates\_VP\_presence\_absence\_model.csv”

There’s also this : “results/parameter\_estimates\_VP\_R2T\_Beta.csv” : it looks at the beta estimates for each level of the environmental factor variables, but it didn’t work (NAs). Not sure why but doesn’t matter too much for us I think.

* Beta plot : the posterior beta estimates for each plant \* covariate. Corresponding values in “results/parameter\_estimates\_Beta\_ presence-absence model.xls”
* Omega plots : associations between species at the site and id level.

## Evaluate model fit

I’m not sure I did this right.

nfolds = NULL #Default: two-fold cross-validation. There’s a place in the code where I think we’re supposed to specify the column (variable) over which the fold should be done. I didn’t..

It’s all stored in the “models/MF\_thin\_1000\_samples\_250\_chains\_4\_nfolds\_2.RData”, and the results are in “results/model\_fit\_nfolds\_2.pdf”. But I don’t understand these plots.

## Make predictitons

See the “results/predictions.pdf”.