HMSC 1st model runs - KERGUELEN

The purpose here is to run a first series of models for Kerguelen 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 ra
# 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     1777.4  3.241821             K10A02   23   12  2010    22         EE  1
## 2     1777.4  3.241528             K10A03   23   12  2010    22          E  2
## 3     1777.4  3.241357             K10A04   23   12  2010    24          E  3
## 4     1761.6  3.195654             K10A05   23   12  2010    21         EE  4
## 5     1761.6  3.182520             K10A06   23   12  2010    21          E  5
## 6     1761.6  3.159375             K10A07   23   12  2010    21          E  6
```

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

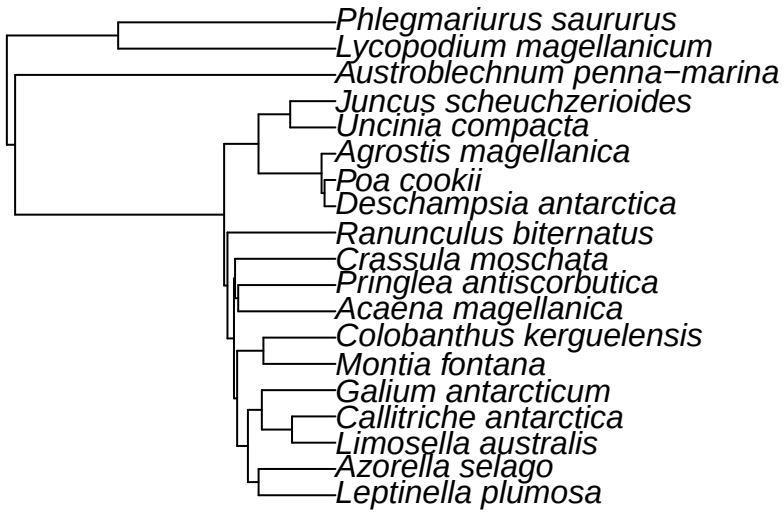
mean_temp = mean temperature downscaled to ~30m, based on mean annual daily mean air temperatures averaged over 1 year (bio1 CHELSA), from 1981-2010

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 Notogrammitis crassior for Kerguelen
phylo_ker <- ape::drop.tip(phylotree, tip = "Notogrammitis crassior")
plot(phylo_ker)
```



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/ 2450 levels "K10A02","K10A03",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ s            : NULL
##  $ sDim         : num 0
##  $ spatialMethod: NULL
##  $ x            : NULL
##  $ xDim         : num 0
##  $ N            : int 2450
##  $ 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/ 2450 levels "K10A02","K10A03",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ s            : NULL
##  $ sDim         : num 0
##  $ spatialMethod: NULL
##  $ x            : NULL
##  $ xDim         : num 0
##  $ N            : int 2450
##  $ 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_ker,
         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 "ker/results/parameter_estimates_ex2.pdf").

- Variance partitioning plot: The corresponding values are in "ker/results/parameter_estimates_VP_presence_absence_

There's also this : "ker/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 "ker/models/MF_thin_1000_samples_250_chains_4_nfolds_2.RData", and the results are in "ker/results/model_fit_nfolds_2.pdf". But I don't understand these plots.

Make predictitons

See the "ker/results/predictions.pdf".