

Simulation Framework

May 2024

Introduction

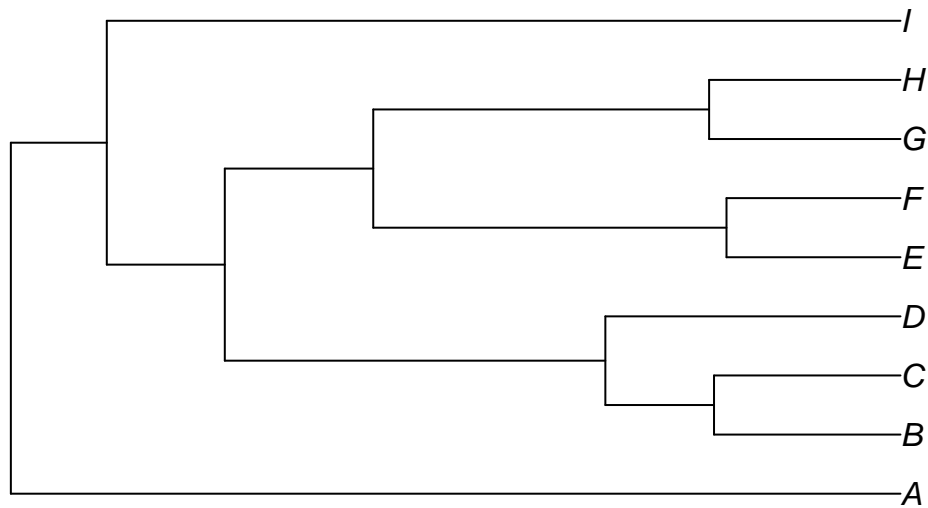
Here we describe the simulation framework used to test and analyse our modeling framework. We utilize functions from library **ape** and **phytools** to complete this simulation.

The simulation framework begins with generating a phylogenetic tree. We provide code below. Here, we call the phylogenetic tree used for our simulation experiments.

1. Simulate phylogeny

```
## Loading required package: maps
```

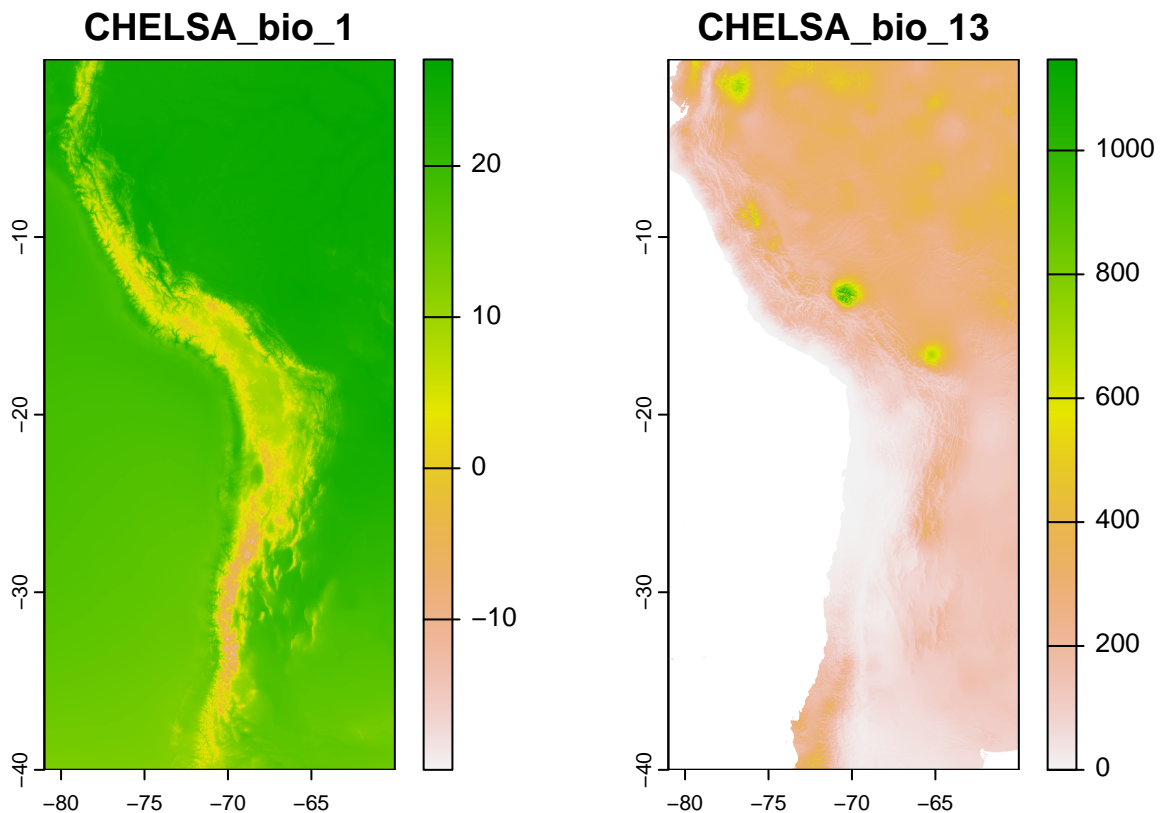
```
tree_time <- 8  
tr <- ape::rbdtree(0.1, 0, Tmax = tree_time)  
tr$edge.length <- tr$edge.length / max(phytools::nodeHeights(tr)[,2]) * 1
```



2. Set up a spatial grid

The spatial grid is set up systematically to sample points from the rasters below.

```
GRID <- createGrid(  
  USER_GRID,  
  nsamples = 100, USER_PATH,  
  xmin = -10, xmax = 10,  
  ymin = -10, ymax = 10  
)  
  
env <- wrangleEnv(epath1 = "~/env", grid1 = GRID)  
env_files <- file.path("~/env", c("CHELSA_bio_1.tif", "CHELSA_bio_13.tif"))  
ras <- terra::rast(env_files)  
extent = c(-81, -60, -40, 0)  
ras <- terra::crop(ras, terra::ext(extent))  
terra::plot(ras)
```



3. Simulate species-environment coefficients using an evolutionary process

For a specific set of parameters and a given tree, the species-environment coefficients (niche parameters) are estimated using either the Ornstein-Uhlenbeck (OU) Process or the Brownian Motion (BM) Process.

```

TREE <- tree
NTRAITS <- 2
SIGMA2 <- 1
ALPHA <- 1
  if (ALPHA == 0) {
    bb <- phytools::fastBM(TREE, sig2 = SIGMA2, a0 = 1, nsim = NTRAITS)
  } else {
    bb <- phytools::fastBM(TREE,
      alpha = ALPHA, theta = 1,
      sigma2 = SIGMA2, nsim = NTRAITS
    )
  }
  colnames(bb) <- paste0("trait-", seq_len(ncol(bb)))

```

4. Simulate species' abundance or presence/absence

Using the environmental variable, the spatial grid and the niche parameters, we can now simulate a species' abundance or presence/absence. We then split the species' data into train and test sets (where 70% of the data is training data and 30% is held-out for validation). Here, we show an example dataset for presence-absence data.

```

# Simulates an OU/BM process
all_data <- simulateAbundanceData_new(
  grid = GRID,
  nsamples = NSAMPLES,
  binary = BINARY,
  print = TRUE,
  tree,
  vars = c("CHELSA_bio_1.tif", "CHELSA_bio_13.tif"),
  ntraits = NTRAITS,
  sigma2 = SIGMA2,
  alpha = ALPHA,
  epath = EPATH
)

```

```

## For Y_train[, sp] specs are as follows: Dimensions are 70, 9
## and number of presence points per species are 34, 38, 37, 41,
## 39, 37, 35, 38, 38.
##
## For Y_test[, sp] specs are as follows: Dimensions are 30, 9 and
## number of presence points per species are 14, 15, 13, 14, 13,
## 13, 14, 14, 14.

##      A B C D E F G H I
## 512 1 1 1 1 1 1 1 0 0
## 952 0 0 0 0 0 0 0 0 0
## 381 1 0 1 0 0 1 0 0 1
## 864 1 1 1 1 0 1 1 1 1
## 738 1 0 0 0 0 0 0 1 1
## 335 0 0 0 0 0 0 0 0 0

```

This code and all required functions are available at the associated GitHub Repository.