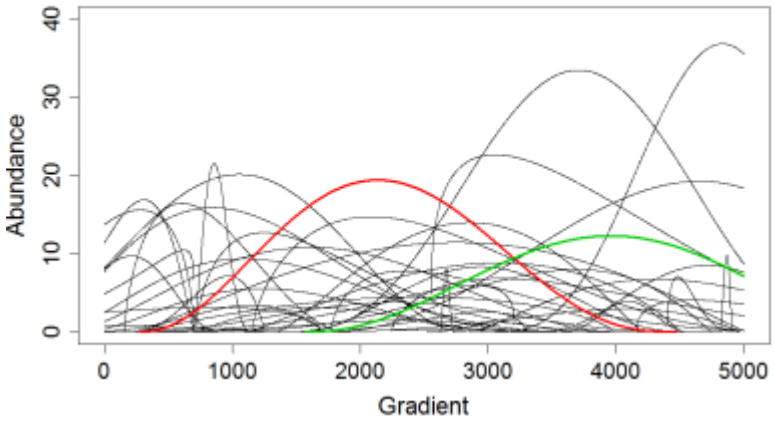
# Simulated ecological data

## Source of data

Zelený D. (unpubl.), script is based on the simulation model written by Fridley et al. (2007) (see Appendix S2 of their paper), which itself was based on the work of Minchin (1987).

## Description of the dataset




These simulated community datasets represent the model of community, which is fully based on the ecological niche theory. Unimodal species response curves are randomly distributed along one (or two, respectively) virtual ecological gradients, reflecting the probability of species occurrence in given part of the gradient (species response curve is based on Beta function). Each species is defined by its ecological optimum along the gradient, niche width, maximum probability of occurrence and few other parameters. In the next step, random positions along gradient are generated, and within each position ("sample") are collected individual species in the following way: first, random number is generated, corresponding to the number of individuals in a given sample; than, each individual is randomly assigned to a species and probability of the assignment to given species is weighted by probability of this species occurrence in particular part of the gradient. One species could be hence assigned to more individuals per sample, if its probability of occurrence in given part of the gradient is high. In case of two virtual gradients, the probability of occurrence for particular species is given by multiplying the probabilities of given species along each of the gradient. For details, see the scripts below.

## Parameters of the files

- simul1 - 1 gradient (length 5000 units), 500 samples, 300 species[1)]
- simul2 - 2 gradients of different length (5000 a 2000 units), 500 samples, 300 species
- simul3 - 2 gradients of the same length (5000 units), 500 samples, 300 species
- simul.short - 2 gradients of different length, both rather short (1100 and 800 units), 70 samples, 300 species, samples are distributed evenly along the gradient (distances between samples along each gradient are exactly 100 units)
- simul.long - 2 gradients of different length, both rather long (5500 and 4000 units), 70 samples, 300 species, samples are distributed evenly along the gradient (distances between samples along each gradient are exactly 100 units)

Note: the number of species (e.g. 300) is a parameter set up for simulated models - the number of species in

the resulting community matrix does not have to fit to number of simulated species, because some of the less abundant species were not “sampled”.

## Environmental variables

| Name of variable | Description |
|---|---|
| gradient | position of the sample along the virtual gradient (for datasets with only one gradient) |
| gradient1, gradient2 | position of each sample along the first and second virtual gradient (for datasets with two gradients) |
| group | classification of samples by modified Twinspan into four groups |

## Species attributes

| | |
|---|---|
| optimum | position of species optima along the virtual gradient (for datasets with only one virtual gradient) |
| niche.width | width of the species niche in units of the virtual gradient (for dataset with only one virtual gradient) |
| optimum1, optimum2 | position of species optima along the first and second virtual gradient (for datasets with two virtual gradients) |
| niche.width1, niche.widht2 | width of the species niche along the first and second virtual gradient (for datasets with two virtual gradients) |

## Data for download

Files containing `-spe` in the name represent presence-absence matrix of species data, files with `-env` contain position of simulated sample along virtual gradient (analogy to measured environmental variable), files with `-specvalues` contain information about position of species optima along the gradient and niche width (both in arbitrary gradient units).

| File name | File type | Description |
|---|---|---|
| simul1-spe.txt | tab-delimited txt format | Sample × species matrix (500 samples in rows, 296 species in columns) |
| simul1-env.txt | tab-delimited txt format | Environmental variable matrix (samples in rows, variables in columns) |
| simul1-specvalues.txt | tab-delimited txt format | Species attribute matrix (species in rows, attributes in columns) |
| simul2-spe.txt | tab-delimited txt format | Sample × species matrix (500 samples in rows, 282 species in columns) |
| simul2-env.txt | tab-delimited txt format | Environmental variable matrix (samples in rows, variables in columns) |
| simul2-specvalues.txt | tab-delimited txt format | Species attribute matrix (species in rows, attributes in columns) |
| simul3-spe.txt | tab-delimited txt format | Sample × species matrix (500 samples in rows, 279 species in columns) |
| simul3-env.txt | tab-delimited txt format | Environmental variable matrix (samples in rows, variables in columns) |
| simul3-specvalues.txt | tab-delimited txt format | Species attribute matrix (species in rows, attributes in columns) |
| simul.short-spe.txt | tab-delimited txt format | Sample × species matrix (70 samples in rows, 300 species in columns) |
| simul.short-env.txt | tab-delimited txt format | Environmental variable matrix (samples in rows, variables in columns) |
| simul.long-spe.txt | tab-delimited txt format | Sample × species matrix (70 samples in rows, 300 species in columns) |
| simul.long-env.txt | tab-delimited txt format | Environmental variable matrix (samples in rows, variables in columns) |

## Script for direct import of data to R

```
simul1.spe <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul1-spe.txt
', row.names = 1)
simul1.env <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul1-env.txt
', row.names = 1)
simul1.specvalues <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul1-specval
ues.txt', row.names = 1)

simul2.spe <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul2-spe.txt
', row.names = 1)
simul2.env <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul2-env.txt
', row.names = 1)
simul2.specvalues <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul2-specval
ues.txt', row.names = 1)

simul3.spe <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul3-spe.txt
', row.names = 1)
simul3.env <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul3-env.txt
', row.names = 1)
simul3.specvalues <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul3-specval
ues.txt', row.names = 1)

simul.short.spe <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul.short-sp
e.txt', row.names = 1)
simul.short.env <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul.short-en
v.txt', row.names = 1)

simul.long.spe <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul.long-spe
.txt', row.names = 1)
simul.long.env <- read.delim
('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/simul.long-env
.txt', row.names = 1)
```

## Scripts for creating simulated datasets

- simulated ecological data along one virtual gradient
- simulated ecological data along two virtual gradients

**Notes:**

- Simulated data along one and two ecological gradients, respectively, can be now prepared faster using functions `simul.comm` and `simul.comm.2` from the package `simcom`.
- Function `compas` in package `CommEcol` written by Adriano Sanches Melo should be a direct implementation of Minchin's software COMPAS (Minchin 1987). The principles are similar (actually, Fridley et al. 2007 paper cites Minchin's paper using COMPAS), but `compas` allows generation of community matrix in more than two dimensions and adding quantitative and qualitative noise.
- Even more comprehensive is package `coenocliner` developed by Gavin Simpson - apart from Minchin's model, it can simulate bunch of other types of community data along coenocline.

## References


- Fridley J.D., Vandermast D.B., Kuppinger D.M., Manthey M. & Peet, R.K. (2007): Co-occurrence-based assessment of habitat generalists and specialists: a new approach for the measurement of niche width. *Journal of Ecology* 95: 707-722 pdf Appendix S2
- Minchin P.R. (1987): Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71: 145-156.


[1)]
Number of species in resulting community matrix is slightly smaller - this is because 300 is the number of species used in simulated community, but not all species made it through random sampling into realized community matrix.


From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
**https://anadat-r.davidzeleny.net/doku.php/en:data:simul**

Last update: **2021/03/25 08:10**