

Practical: Using simmr

Andrew Parnell

Introduction

Welcome to the practical! We will learn about:

- Loading data into simmr
- Running simmr
- Getting output from simmr
- Using simmr for single observations
- Setting prior distributions on the dietary proportions
- Customising your own simmr output

It's assumed that you already have `simmr` installed via:

```
library(devtools)
install_github('andrewcparnell/simmr')
```

You should then be able to run

```
library(simmr)
```

without error.

This document is a slightly friendlier version of the simmr vignette.

You should follow and run the commands shown in the grey boxes below. At various points you will see a horizontal line in the text which indicates a question you should try to answer, like this:

What words does the following command print to the console?

```
print("Hello World")
```

If you get stuck, please get our attention and we will try to help! There are no prepared answers to these questions so keep your own record as you go. At the end of the practical are harder questions which you can attempt if you get through all of the material. If you find any mistakes in the document please let us know.

You can run the code from these practicals by loading up the `.Rmd` file in the same directory in Rstudio. This is an R markdown document containing all the text. Feel free to add in your own answers, or edit the text to give yourself extra notes. You can also run the code directly by highlighting the relevant code and clicking Run.

Loading data into simmr

Unlike SIAR and MixSIAR which have both menu systems and script-based ways of interacting with the software, simmr only allows a script-based version. This makes it slightly harder to use for those unfamiliar with R but allows for much more flexibility in the long run. We will focus on using the script-based way of running simmr.

The key differences between simmr and other SIMMs are:

- simmr has a slightly richer mixing model than SIAR, based on code from the Parnell et al 2013 Environmetrics paper

- simmr uses ggplot2 to create graphs and JAGS to run the mixing model
- simmr has simple functions to load data, create iso-space plots, and summarise/plot output
- simmr has some simple functions for comparing and combining sources, and for comparing groups

We will use the data included with SIAR (make sure you have installed SIAR using the instructions in the ‘using SIAR’ practical) to run simmr. When running your own models you should try to keep your data in the same format as these examples. At minimum you need two files to get simmr working; a consumers file and a sources file. The simplest geese data set is obtained via:

```
library(siar)
data(geese1demo)
print(geese1demo)
```

This data set has two columns, one for each isotope, and 9 individuals. A useful command for learning about the structure of an R data set is `str`, especially for large data objects:

```
str(geese1demo)

##  num [1:9, 1:2] 10.2 10.4 10.4 10.5 10.2 ...
## - attr(*, "dimnames")=List of 2
##  ..$ : NULL
##  ..$ : chr [1:2] "d15NP1" "d13CP1"
```

We can see that it has 9 rows, 2 columns, and is of numeric (`num`) mode. The two column labels refer to the $\delta^{15}\text{N}$ and $\delta^{13}\text{C}$ isotope values.

The sources data can be obtained from:

```
data(sourcesdemo)
print(sourcesdemo)

##      Sources Meand15N  SDd15N Meand13C  SDd13C
## 1    Zostera  6.488984 1.4594632 -11.17023 1.2149562
## 2     Grass  4.432160 2.2680709 -30.87984 0.6413182
## 3  U.lactuca 11.192613 1.1124385 -11.17090 1.9593306
## 4 Enteromorpha 9.816280 0.8271039 -14.05701 1.1724677
```

We can see that there are 4 sources, with their names in the first column. The remaining columns refer to the means and standard deviations for each source on each isotope. The isotopes need to be in the same order as the consumer data in `geese1demo`. Note the structure of this object

```
str(sourcesdemo)

## 'data.frame': 4 obs. of 5 variables:
## $ Sources : Factor w/ 4 levels "Enteromorpha",...: 4 2 3 1
## $ Meand15N: num 6.49 4.43 11.19 9.82
## $ SDd15N : num 1.459 2.268 1.112 0.827
## $ Meand13C: num -11.2 -30.9 -11.2 -14.1
## $ SDd13C : num 1.215 0.641 1.959 1.172
```

It’s a data frame. This is an R data type which can store both text and numbers, useful for storing the source names as well as their isotope values.

We could run simmr with just these two data files. However, this would produce a pretty poor model as we don’t have any corrections for trophic enrichment factors (TEFs). The TEFs file looks just like the source file:

```
data(correctionsdemo)
print(correctionsdemo)

##      Source Mean15N Sd15N Mean13C Sd13C
## 1    Zostera   3.54  0.74   1.63  0.63
```

```
## 2      Grass      3.54 0.74    1.63 0.63
## 3    U.lactuca    3.54 0.74    1.63 0.63
## 4 Enteromorpha   3.54 0.74    1.63 0.63
```

Lastly, if we have concentration dependence we can include these too. The data for the Geese comes from:

```
data(concdepdemo)
print(concdepdemo)
```

```
##      Sources Meand15N SDd15N Meand13C SDd13C
## 1    Zostera  0.0297 0.0097   0.3593 0.0561
## 2     Grass  0.0355 0.0063   0.4026 0.0380
## 3    U.lactuca 0.0192 0.0053   0.2098 0.0327
## 4 Enteromorpha 0.0139 0.0057   0.1844 0.1131
```

Note that although this data set includes standard deviations on the concentration dependencies, they are currently not used by `simmr` in the model run.

If you were loading these data sets in yourself, it's best to store them in the same directory and then load them in from there, e.g.:

```
# Set the working directory (where R looks first for files)
setwd('path/to/files')
# Read in consumers
consumers = read.table('my_consumer_file.txt',header=TRUE)
# Read in sources
sources = read.table('my_sources_file.txt',header=TRUE)
# Read in TEFs
TEFs = read.table('my_TEF_file.txt',header=TRUE)
```

The extra `header=TRUE` argument tells R that there are column names at the top of the file.

-
1. What is the structure of the TEFs object? How many rows and columns does it have?
 2. There's another data object that comes with `siar` called `geese2demo`. How many rows and columns does this have?
 3. Create some simple scatter plots of the `geese1demo` data using `plot`. See if you can add in the source means corrected for the TEF means (hint: add the means together and then plot using `points`)
-

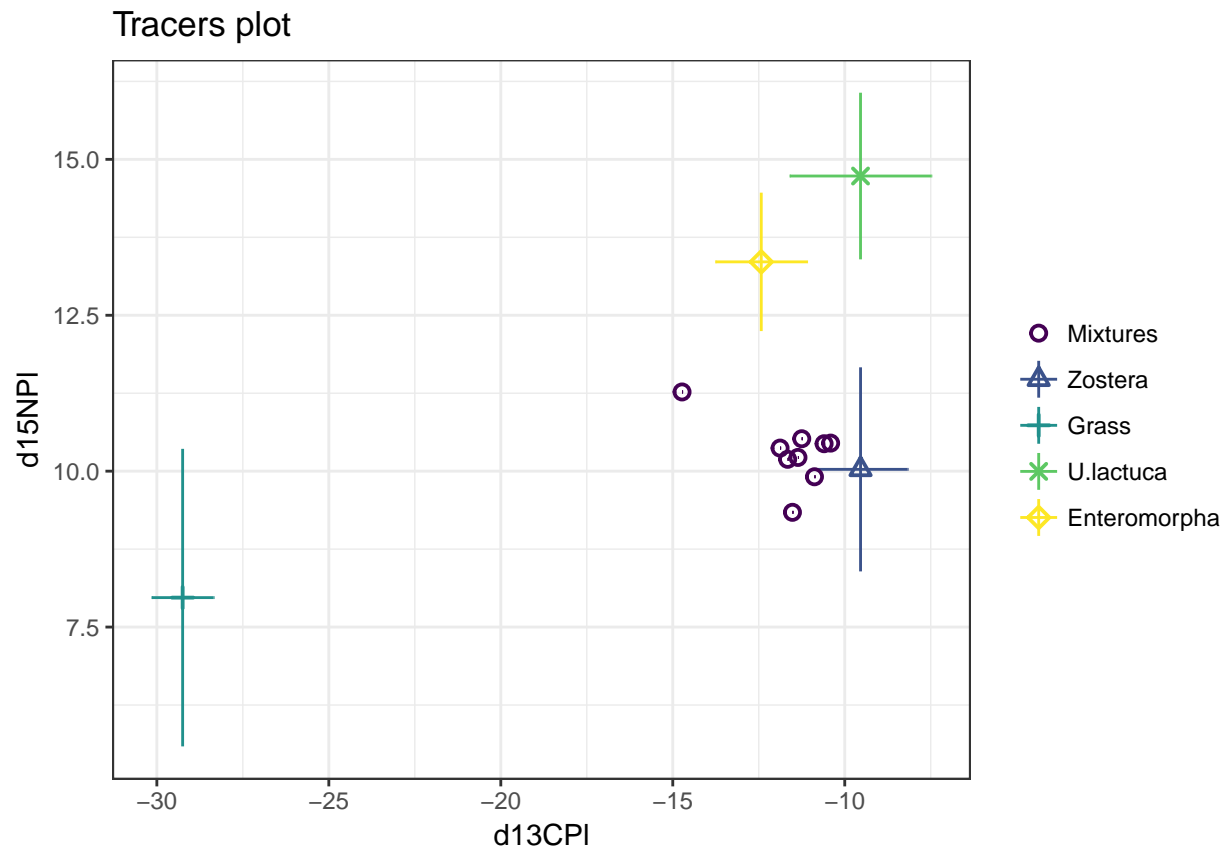
Loading data into `simmr`

We are now in a position to run `simmr`. The first function to use is `simmr_load`. You can find help on this function by typing the name with a `?` in front. If you are using Rstudio you can use the `<TAB>` key to complete your command once you have typed in the first few characters. To load the Geese data into `simmr`, type:

```
simmr_in = simmr_load(mixtures = geese1demo[,c(2,1)],
                      source_names = as.character(sourcesdemo[,1]),
                      source_means = sourcesdemo[,c(4,2)],
                      source_sds = sourcesdemo[,c(5,3)],
                      correction_means = correctionsdemo[,c(4,2)],
                      correction_sds = correctionsdemo[,c(5,3)],
                      concentration_means = concdepdemo[,c(4,2)])
```

Once loaded we can create an isospace plot very simply with:

```
plot(simmr_in)
```



You should see that the consumers are inside the mixing polygon (or *convex hull*) of the sources. The consumers are close to the Zostera source, so we would expect this to come out as the main dietary proportion.

We can also print out the details of the current model:

```
print(simmr_in)
```

```
## This is a valid simmr input object with 9 observations, 2 tracers, and 4 sources.
## The source names are: Zostera, Grass, U.lactuca, Enteromorpha.
## The tracer names are: d13CPI, d15NPI.
```

Running simmr

To run simmr we use the `simmr_mcmc` function:

```
simmr_out = simmr_mcmc(simmr_in)
```

simmr now runs the MCMC algorithm (using JAGS) and, whilst running, reports the percentage complete. When finished, the first thing to do is to check convergence

```
summary(simmr_out, type = 'diagnostics')
```

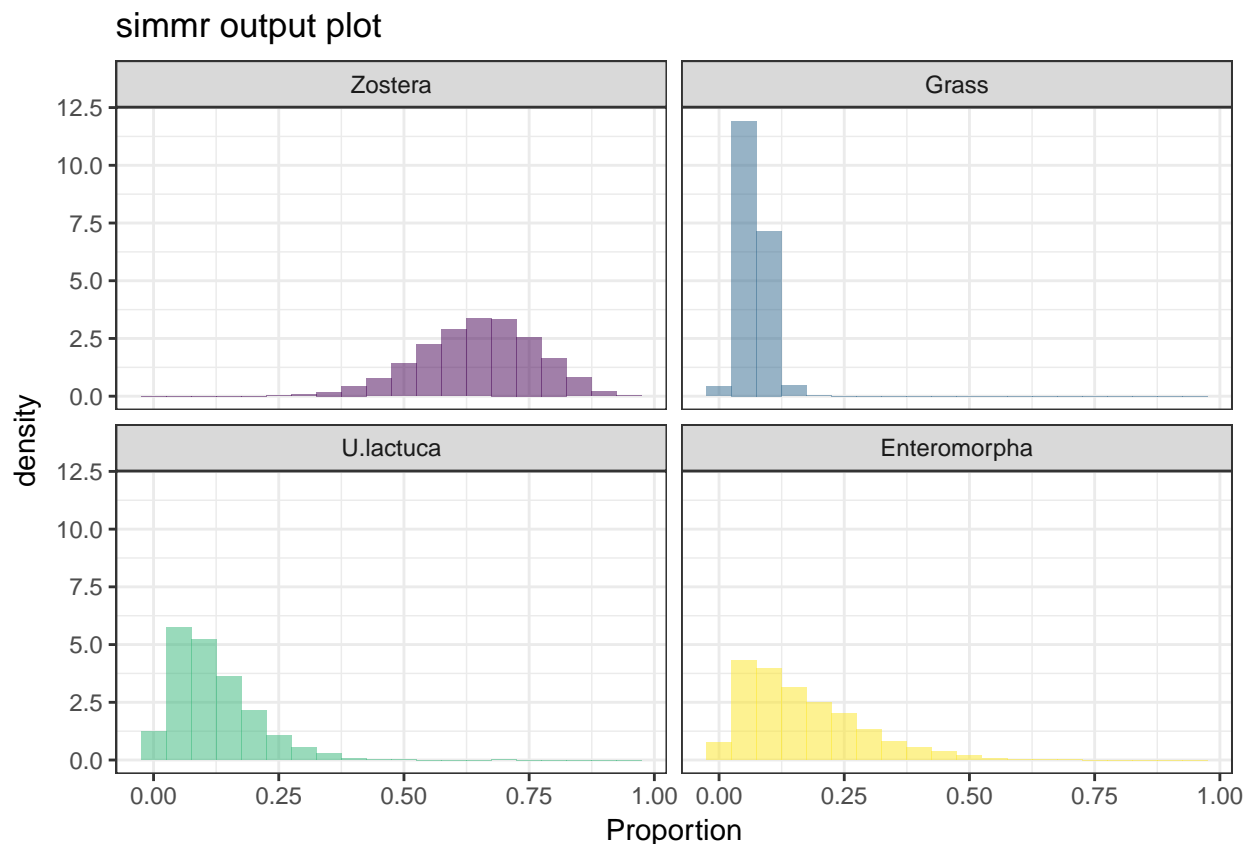
```
##
## Summary for 1
## Gelman diagnostics - these values should all be close to 1.
## If not, try a longer run of simmr_mcmc.
```

```
##          Point est. Upper C.I.
## Zostera          1.01      1.02
## Grass            1.01      1.04
## U.lactuca        1.01      1.02
## Enteromorpha     1.03      1.08
## sd_d13CP1        1.00      1.01
## sd_d15NP1        1.01      1.01
```

These are the Brooks-Gelman-Rubin diagnostics discussed in the lecture notes. All values are close to 1 so the model has achieved satisfactory convergence. The rule of thumb is usually that the model has converged ok if the values are less than 1.1.

We can plot the posterior distributions of the dietary proportions with:

```
plot(simmr_out, type = 'histogram')
```



If you want more textual output you can get it with:

```
summary(simmr_out, type = c('statistics', 'quantiles'))
```

```
##
## Summary for 1
##          2.5%  25%  50%  75% 97.5%
## Zostera    0.405 0.573 0.655 0.729 0.852
## Grass      0.027 0.051 0.067 0.084 0.125
## U.lactuca  0.015 0.058 0.102 0.161 0.312
## Enteromorpha 0.021 0.074 0.139 0.234 0.451
## sd_d13CP1  0.035 0.293 0.608 1.059 2.219
## sd_d15NP1  0.021 0.193 0.404 0.698 1.578
```

```
##           mean    sd
## Zostera    0.648 0.116
## Grass      0.069 0.025
## U.lactuca  0.118 0.079
## Enteromorpha 0.165 0.115
## sd_d13CP1  0.744 0.601
## sd_d15NP1  0.505 0.419
```

This will produce a quantiles for the posterior proportions (useful for 50% and 95% uncertainty intervals), and the estimated mean and standard deviation. It will also give the same estimates for the residual standard deviations for each isotope. These will tend to be large when the consumers lie outside the source mixing polygon.

If you want held on these functions use `?summary.simmr_output` and `?plot.simmr_output`

-
1. What's the structure of the `simmr_out` object? Can you see anything you recognise in it? Try accessing different parts of it using the `$` notation, e.g. `simmr_out$input$n_obs`
 2. Try the command `plot(simmr_out, type = 'matrix')`. What does this produce? What other types of plots are available?
 3. Try running the model again without including the `correctionsdemo` data. What happens to the isospace plot?
-

Longer simmr runs

If you want to be really certain of convergence you can run `simmr` for more iterations with some extra arguments. The extra arguments are:

1. `iter` which sets the total number of iterations. The default is 10,000
2. `burn` which sets the number of initial iterations to remove. The default is 1,000
3. `thin` which sets the amount of thinning (removal) of iterations to avoid autocorrelation in the output values. The default is 10, which means `simmr` will keep only every 10th iteration
4. `n.chain` which sets the number of chains (i.e. the number of runs with different starting values). The default is 4 chains.

Usually the default values will be fine, but you could double them if you wanted a longer run. The resulting number of iterations kept by `simmr` for the posterior distribution is `n.chain*(iterations-burnin)/thinby`. It's usually not a good idea to store more than 10,000 iterations unless you have lots of RAM.

To include such extra values a run might be:

```
simmr_out_long = simmr_mcmc(simmr_in,
                           mcmc.control = list(iter = 20000,
                                                burn = 2000,
                                                thin = 20,
                                                n.chain = 4))
```

-
1. Without checking, how many iterations will the command above save?
 2. Did the results change much between the shorter and longer run?
 3. Were the convergence results better for the longer run (i.e. were the BGR diagnostic values closer to 1)?
-

Working with multiple groups

Sometimes you might be interested in running `simmr` for multiple different groups of consumers. These different groups might be different sexes, different sampling periods, different locations, etc. `simmr` will run these simultaneously and store the output for easier plots and comparison.

The data which are included in `simmr` for multiple groups analysis can be found with:

```
data(geese2demo)
head(geese2demo, 15)
```

```
##      Group d15NP1 d13CP1
## [1,]      1  10.22 -11.36
## [2,]      1  10.37 -11.88
## [3,]      1  10.44 -10.60
## [4,]      1  10.52 -11.25
## [5,]      1  10.19 -11.66
## [6,]      1  10.45 -10.41
## [7,]      1   9.91 -10.88
## [8,]      1  11.27 -14.73
## [9,]      1   9.34 -11.52
## [10,]     2  11.68 -15.89
## [11,]     2  12.29 -14.79
## [12,]     2  11.04 -17.64
## [13,]     2  11.46 -16.97
## [14,]     2  11.73 -17.25
## [15,]     2  12.29 -14.77
```

```
str(geese2demo)
```

```
##  num [1:251, 1:3] 1 1 1 1 1 1 1 1 1 2 ...
## - attr(*, "dimnames")=List of 2
##  ..$ : NULL
##  ..$ : chr [1:3] "Group" "d15NP1" "d13CP1"
```

This is a much bigger data set. The first column contains the group number. To run `simmr` you now need to specify the groups so that it can separate out the model. We can see how many and how large the groups are with:

```
table(geese2demo[, 'Group'])
```

```
##
##  1  2  3  4  5  6  7  8
##  9 29 74 10 41 20 32 36
```

so 8 groups ranging from 9 to 74 observations. `simmr` will work with up to 30 groups. There needs to be at least 3 observations per group for `simmr` to run, but really 5 or more is desirable if you want to properly estimate the residual error.

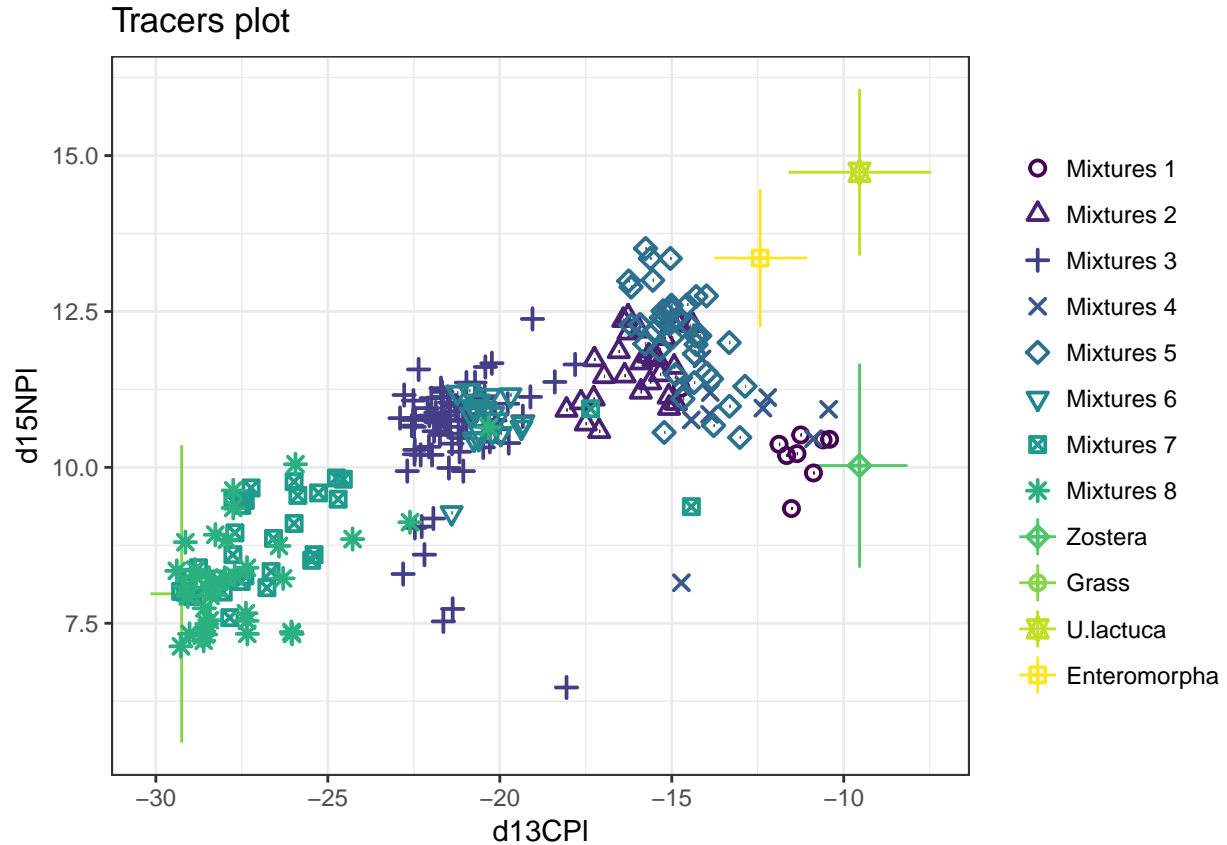
To run `simmr` with this data set it's the same as before:

```
simmr_in2 = simmr_load(mixtures = geese2demo[,c(3,2)],
                        source_names = as.character(sourcesdemo[,1]),
                        source_means = sourcesdemo[,c(4,2)],
                        source_sds = sourcesdemo[,c(5,3)],
                        correction_means = correctionsdemo[,c(4,2)],
                        correction_sds = correctionsdemo[,c(5,3)],
                        concentration_means = concdepdemo[,c(4,2)],
```

```
group = as.integer(geese2demo[,1]))
```

Again we can do an isospace plot with:

```
plot(simmr_in2, group = 1:8)
```



We run simmr the same as before too:

```
simmr_out2 = simmr_mcmc(simmr_in2)
```

The summaries and convergence diagnostics are created with, e.g.:

```
summary(simmr_out2, type = c('diagnostics', 'quantiles'), group = 3)
```

```
##
## Summary for 3
## Gelman diagnostics - these values should all be close to 1.
## If not, try a longer run of simmr_mcmc.
##           Point est. Upper C.I.
## Zostera      1.15      1.42
## Grass        1.45      2.09
## U.lactuca     1.65      2.51
## Enteromorpha 1.68      2.57
## sd_d13CPl    1.00      1.01
## sd_d15NPl    1.00      1.00
##           2.5%  25%  50%  75% 97.5%
## Zostera      0.013 0.026 0.036 0.055 0.114
## Grass        0.364 0.392 0.412 0.429 0.456
```

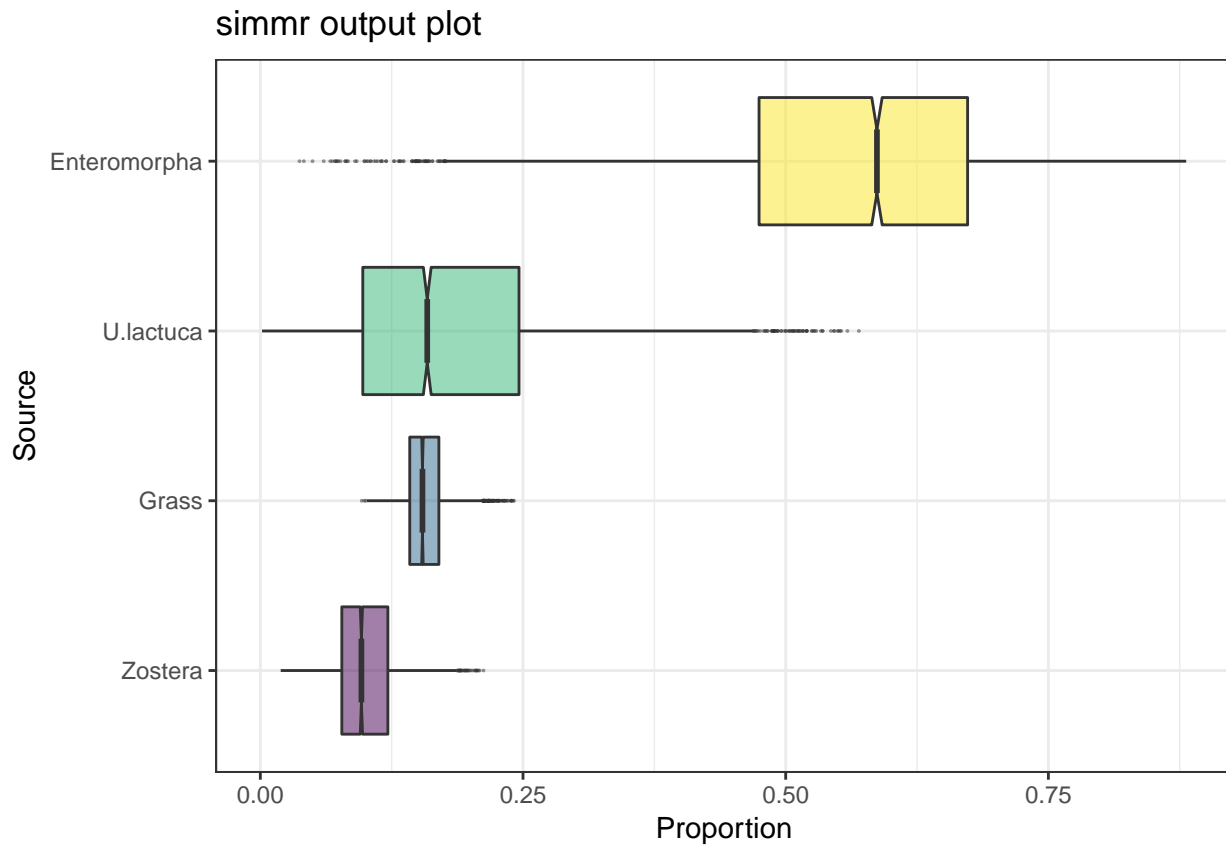


```
## U.lactuca      0.083 0.183 0.291 0.363 0.455
## Enteromorpha  0.070 0.158 0.256 0.377 0.528
## sd_d13CP1     0.014 0.145 0.306 0.501 0.845
## sd_d15NP1     0.008 0.073 0.153 0.265 0.528
```

Running the above without the `group` argument gives the results for all groups (though is pretty long).

We can get some within-group boxplots with:

```
plot(simmr_out2, type = 'boxplot', group = 2)
```

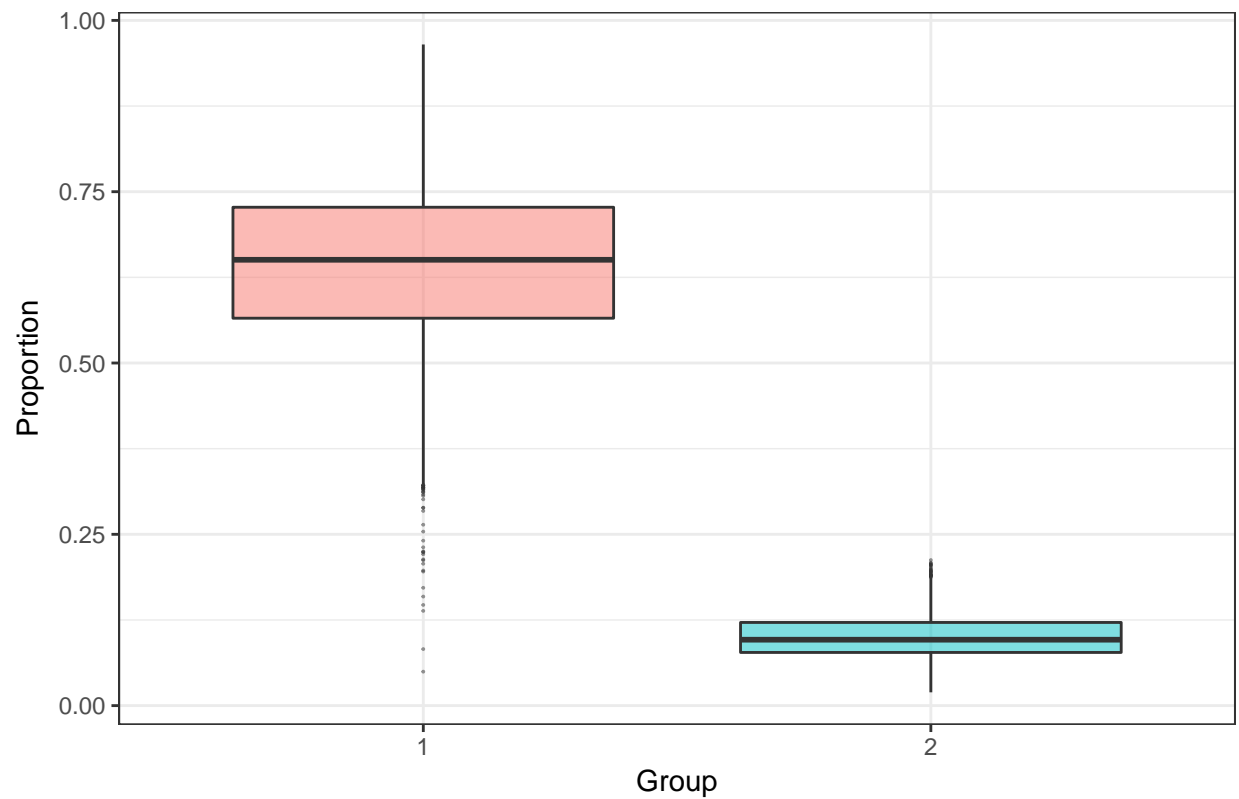


To compare groups we can run, e.g.:

```
compare_groups(simmr_out2, source='Zostera', groups=1:2)
```

```
## Prob ( proportion of Zostera in group 1 > proportion of Zostera in group 2 ) = 1
```

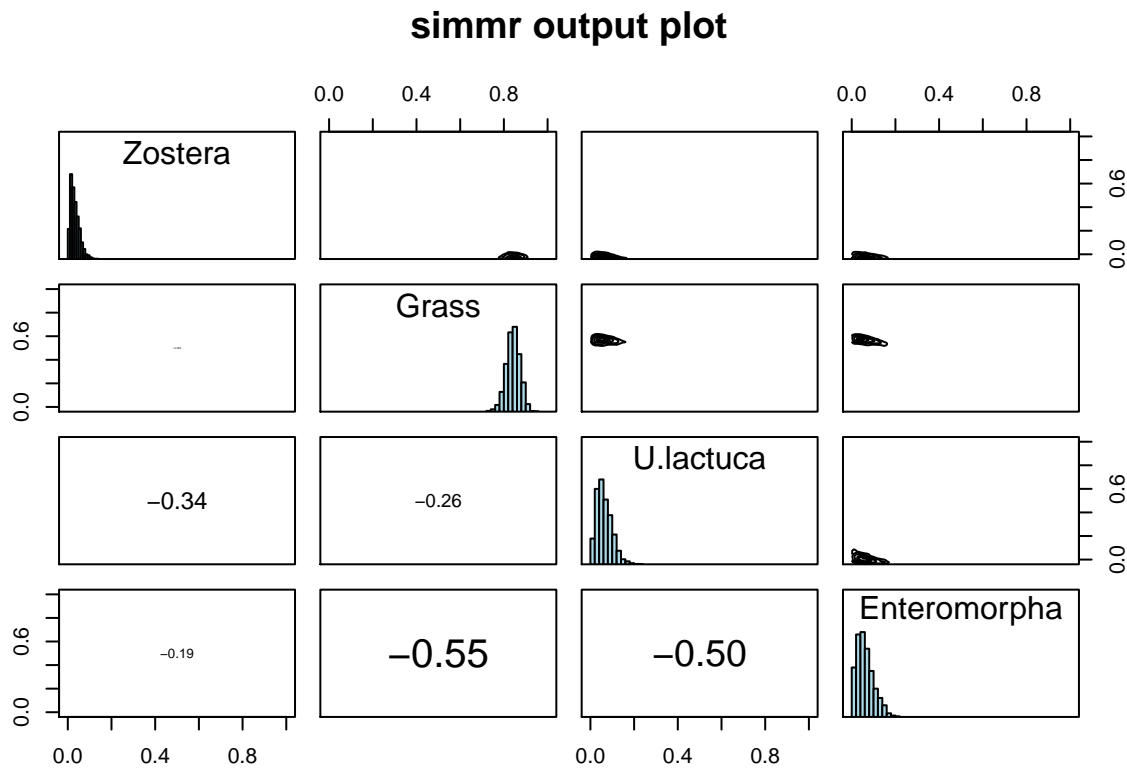
Comparison of dietary proportions for groups 1 and 2 for source Zostera



This will also estimate the probability that this source is being eaten by one group more than the other.

Finally, the matrix plot (discussed in the lectures) can be created with:

```
plot(simmr_out2, type = 'matrix', group = 7)
```



This is a really useful plot as it provides the histograms and the relationships between the sources, potentially identifying which sources are impossible to discern between in the model. It takes a little bit of practice to interpret a matrix plot.

Running the model for individual observations

When you have just a single observation it is impossible to estimate the residual standard deviation. However you can still estimate the dietary proportions and `simmr` will run this as standard when provided with only one observation by forcing the prior distribution on the residual standard deviation to be tiny. We can create a simple example by just taking a single row from the geese data:

```
simmr_in3 = simmr_load(mixtures = geese1demo[1,c(2,1), drop = FALSE],
                      source_names = as.character(sourcesdemo[,1]),
                      source_means = sourcesdemo[,c(4,2)],
                      source_sds = sourcesdemo[,c(5,3)],
                      correction_means = correctionsdemo[,c(4,2)],
                      correction_sds = correctionsdemo[,c(5,3)],
                      concentration_means = concdepdemo[,c(4,2)])
simmr_out3 = simmr_mcmc(simmr_in3)
```

The extra argument `drop = FALSE` is used above because the mixtures have to be in matrix format (without this R will turn the resulting object into a vector rather than a matrix).

```
summary(simmr_out3, type = 'quantiles')
```

```
##
## Summary for 1
```

##	2.5%	25%	50%	75%	97.5%
## Zostera	0.049	0.210	0.369	0.543	0.814
## Grass	0.021	0.067	0.124	0.237	0.603
## U.lactuca	0.020	0.087	0.166	0.302	0.647
## Enteromorpha	0.021	0.091	0.184	0.323	0.654
## sd_d13CP1	0.000	0.000	0.001	0.001	0.001
## sd_d15NP1	0.000	0.000	0.001	0.001	0.001

Note that the estimates of the residual standard deviation are all tiny values.

Adding in your own prior information

Occasionally it is the case that previous studies have given insight into the likely values of the dietary proportions for your study. You can use this external information to guide the model by changing the prior distributions used for the transformed CLR-distribution (see module on: ‘From SIAR to MixSIAR’). If prior information is available, it is usually a good idea to use it, as it means the model will often converge quicker, and yield more realistic results.

`simmr` has a special function for the inclusion of prior distributions. For example suppose we had prior information that the four sources in the Geese example had means (0.5, 0.1, 0.2, 0.2) and standard deviations (0.1, 0.05, 0.01, 0.02), we would run:

```
prior = simmr_elicit(simmr_in,
                    proportion_means = c(0.5,0.1,0.2,0.2),
                    proportion_sds = c(0.1,0.05,0.01,0.02))
```

This requires some optimisation routines in the background and can often be slow. The function will print out some values to use in the prior distribution for the CLR-transformed values. We can use these as follows:

```
simmr_out_prior = simmr_mcmc(simmr_in,
                             prior_control = list(means = prior$mean,
                                                  sd = prior$sd))
```

-
1. Try several different prior estimates and see how they change the posterior dietary proportions with `plot`.
 2. An alternative default prior for `simmr` would be when all of the mean prior proportions are set to 1 divided by the number of sources, and the prior standard deviations are set to be very large. What effect does using this prior have on the different Geese data sets?
-

Creating your own plots and tables

Often what you want to create isn’t exactly part of the `simmr` toolkit. Maybe the plots don’t look right, or maybe you want to compare two different groups in a particular way. To do this, you can get at the `simmr` output yourself, and then play with it as you want.

Whenever `simmr` creates the dietary proportions using e.g. `simmr_mcmc`, it stores the output as an R *list*. You can see everything in the list with:

```
str(simmr_out)

## List of 2
## $ input :List of 13
## ..$ mixtures : num [1:9, 1:2] -11.4 -11.9 -10.6 -11.2 -11.7 ...
```

```
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : NULL
## .. ..$ : chr [1:2] "d13CP1" "d15NP1"
## ..$ source_names : chr [1:4] "Zostera" "Grass" "U.lactuca" "Enteromorpha"
## ..$ source_means : 'data.frame': 4 obs. of 2 variables:
## .. ..$ Meand13C: num [1:4] -11.2 -30.9 -11.2 -14.1
## .. ..$ Meand15N: num [1:4] 6.49 4.43 11.19 9.82
## ..$ source_sds : 'data.frame': 4 obs. of 2 variables:
## .. ..$ SDd13C: num [1:4] 1.215 0.641 1.959 1.172
## .. ..$ SDd15N: num [1:4] 1.459 2.268 1.112 0.827
## ..$ correction_means : 'data.frame': 4 obs. of 2 variables:
## .. ..$ Mean13C: num [1:4] 1.63 1.63 1.63 1.63
## .. ..$ Mean15N: num [1:4] 3.54 3.54 3.54 3.54
## ..$ correction_sds : 'data.frame': 4 obs. of 2 variables:
## .. ..$ Sd13C: num [1:4] 0.63 0.63 0.63 0.63
## .. ..$ Sd15N: num [1:4] 0.74 0.74 0.74 0.74
## ..$ concentration_means: 'data.frame': 4 obs. of 2 variables:
## .. ..$ Meand13C: num [1:4] 0.359 0.403 0.21 0.184
## .. ..$ Meand15N: num [1:4] 0.0297 0.0355 0.0192 0.0139
## ..$ group : Factor w/ 1 level "1": 1 1 1 1 1 1 1 1 1
## ..$ group_int : int [1:9] 1 1 1 1 1 1 1 1 1
## ..$ n_obs : int 9
## ..$ n_tracers : int 2
## ..$ n_sources : int 4
## ..$ n_groups : int 1
## ..- attr(*, "class")= chr "simmr_input"
## $ output:List of 1
## ..$ :List of 4
## .. ..$ : mcmc [1:1000, 1:6] 0.872 0.874 0.801 0.565 0.556 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : NULL
## .. .. ..$ : chr [1:6] "Zostera" "Grass" "U.lactuca" "Enteromorpha" ...
## .. ..- attr(*, "mcpair")= num [1:3] 1010 11000 10
## .. ..$ : mcmc [1:1000, 1:6] 0.454 0.393 0.584 0.375 0.528 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : NULL
## .. .. ..$ : chr [1:6] "Zostera" "Grass" "U.lactuca" "Enteromorpha" ...
## .. ..- attr(*, "mcpair")= num [1:3] 1010 11000 10
## .. ..$ : mcmc [1:1000, 1:6] 0.694 0.594 0.577 0.646 0.492 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : NULL
## .. .. ..$ : chr [1:6] "Zostera" "Grass" "U.lactuca" "Enteromorpha" ...
## .. ..- attr(*, "mcpair")= num [1:3] 1010 11000 10
## .. ..$ : mcmc [1:1000, 1:6] 0.694 0.655 0.728 0.599 0.554 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : NULL
## .. .. ..$ : chr [1:6] "Zostera" "Grass" "U.lactuca" "Enteromorpha" ...
## .. ..- attr(*, "mcpair")= num [1:3] 1010 11000 10
## .. ..- attr(*, "class")= chr "mcmc.list"
## - attr(*, "class")= chr "simmr_output"
```

This will provide quite a lot of output, but the most important part is the element named `output` which contains all of the posterior samples. You can see the first few samples with:

```
head(simmr_out$output[[1]][[1]])
```

```
## Markov Chain Monte Carlo (MCMC) output:
```

```
## Start = 1010
```

```
## End = 1070
```

```
## Thinning interval = 10
```

```
##           Zostera      Grass  U.lactuca Enteromorpha  sd_d13CP1  sd_d15NP1
## [1,] 0.8718384 0.04827234 0.01392824  0.06596098 0.32177951 0.16151918
## [2,] 0.8735883 0.03346297 0.07060161  0.02234708 1.61714638 0.67296844
## [3,] 0.8011074 0.06390687 0.10871326  0.02627246 0.01088534 0.14028942
## [4,] 0.5647788 0.06241410 0.12280161  0.25000544 0.53823326 0.05260842
## [5,] 0.5560245 0.04218585 0.09387139  0.30791828 0.25207949 0.30881025
## [6,] 0.5736655 0.05808635 0.11090812  0.25734008 0.09683072 0.27828306
## [7,] 0.5994426 0.07802773 0.15403538  0.16849432 0.01745416 0.52074337
```

You will see that, for each row, each of the four sources sum to 1:

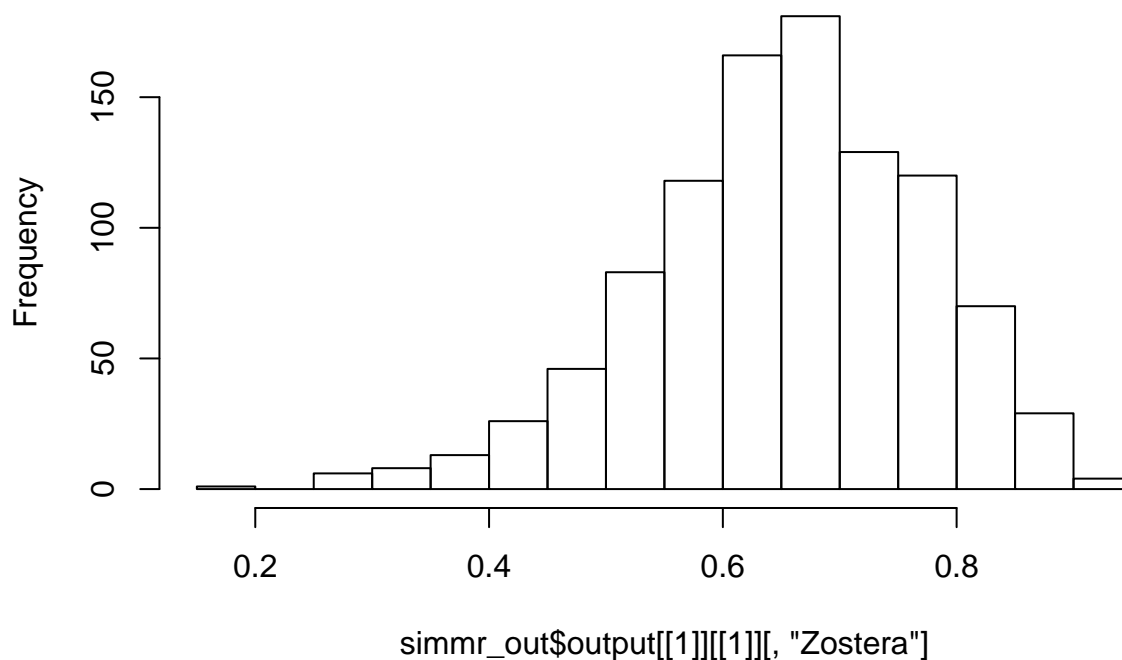
```
sum(simmr_out$output[[1]][[1]][1,1:4])
```

```
## [1] 1
```

You can now create any further that you wish, for example a simple histogram of the posterior proportion for the first chain of *Zostera*:

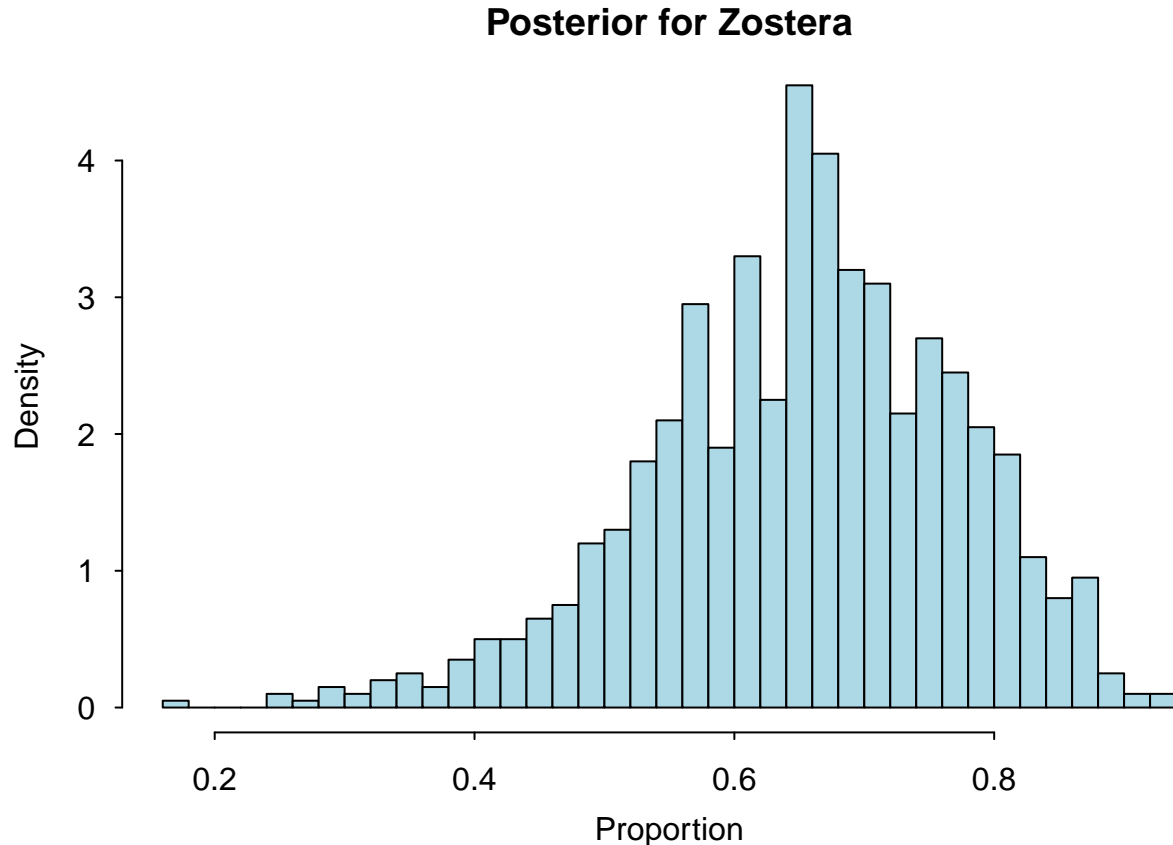
```
hist(simmr_out$output[[1]][[1]][, 'Zostera'])
```

Histogram of `simmr_out$output[[1]][[1]][, "Zostera"]`



This is a bit crude, but with some extra options, you can make this look quite neat:

```
# Set some better options for graphs
par(mar=c(3,3,2,1), mgp=c(2,.7,0), tck=-.01, las=1)
hist(simmr_out$output[[1]][[1]][, 'Zostera'], freq=FALSE, main='Posterior for Zostera',
     xlab='Proportion', col='lightblue', breaks=30)
```



You can also create your own output analysis. For example, what is the 90% credible interval for Grass?

```
quantile(simmr_out$output[[1]][[1]][, 'Grass'], probs=c(0.05, 0.95))
```

```
##          5%          95%
## 0.0328036 0.1131987
```

What is the probability that the consumers ate more Ulva Lactuca than Enteromorpha?

```
sum(simmr_out$output[[1]][[1]][, 'U.lactuca'] > simmr_out$output[[1]][[1]][, 'Enteromorpha']) / nrow(simmr_out)

## [1] 0.462
```

The above counts the number of rows (i.e. iterations) in the output where Ulva Lactuca is higher than Enteromorpha and divides this by the total number of rows. This is essentially what the `compare_groups` function is doing. There is also a `compare_sources` function which also creates something similar to the above.

Finally, if you want to see what simmr is doing behind the scenes, simply type the name of the function without brackets, for example

```
simmr_mcmc
```

If the simmr plot or table doesn't exactly match what you want you can create your own function based on the original one which includes everything you need.

There is even more detail on the extra functions in `simmr` in the `simmr` vignette so make sure to read that before trying to create something complicated yourself.

-
1. Try accessing the output from the second Geese data set (stored above in `simmr_out2`). Try to re-create the above histograms for some of the groups.
 2. Continuing the above, try and calculate the probability that one group ate more of a certain source than another.
-

Some extra tasks

If you finish all the above and want some further tasks to complete try these.

-
1. See if you can re-create the iso-space plot from the raw data from scratch. Refer back to the code in `plot.simmr_input` if you need to.
 2. Try and write your own function to process the output from a `simmr` model run. What would you like to include? Below is a function which just lists the first 15 iterations. You could create something far richer, including means (via `mean`), credible intervals (via `quantile`), correlations (via `cor`) or plots. The `apply` function is often useful here as it will run a function over the rows or columns of a matrix.

```
my_summary = function(x) {
  head(x$output[[1]][[1]],15)
}
my_summary(simmr_out)
```

```
## Markov Chain Monte Carlo (MCMC) output:
```

```
## Start = 1010
```

```
## End = 1160
```

```
## Thinning interval = 10
```

```
##           Zostera      Grass  U.lactuca Enteromorpha  sd_d13CP1  sd_d15NP1
## [1,] 0.8718384 0.04827234 0.01392824  0.06596098 0.32177951 0.16151918
## [2,] 0.8735883 0.03346297 0.07060161  0.02234708 1.61714638 0.67296844
## [3,] 0.8011074 0.06390687 0.10871326  0.02627246 0.01088534 0.14028942
## [4,] 0.5647788 0.06241410 0.12280161  0.25000544 0.53823326 0.05260842
## [5,] 0.5560245 0.04218585 0.09387139  0.30791828 0.25207949 0.30881025
## [6,] 0.5736655 0.05808635 0.11090812  0.25734008 0.09683072 0.27828306
## [7,] 0.5994426 0.07802773 0.15403538  0.16849432 0.01745416 0.52074337
## [8,] 0.6710954 0.03513815 0.12894411  0.16482230 1.32700540 0.08023167
## [9,] 0.8936573 0.04502212 0.02194748  0.03937315 1.18995644 0.90376644
## [10,] 0.8031976 0.04954882 0.09549148  0.05176215 1.06885780 0.97352811
## [11,] 0.8143455 0.05907318 0.06510099  0.06148036 0.30585420 0.65810921
## [12,] 0.7068110 0.07285429 0.09841728  0.12191738 0.31043936 0.38353278
## [13,] 0.6094558 0.08577925 0.20627369  0.09849121 1.53177406 0.70662734
## [14,] 0.7036147 0.06166542 0.17581529  0.05890455 0.42649092 0.15994753
## [15,] 0.5375126 0.07482096 0.13541524  0.25225123 0.54808213 0.07143673
## [16,] 0.7587900 0.11293862 0.10319502  0.02507636 1.34457881 0.82529634
```
