IsotopeR 0.5

John "Jack" Hopkins jbhopkins3@gmail.com; http://jackhopkinswildlife.com/
Jake Ferguson troutinthemilk@gmail.com

November 19, 2015

Estimated Mixing Space

Contents

1　Summary

IsotopeR is a hierarchical stable isotope mixing model used to estimate dietary parameters at the population-, group-, and individual-level. The IsotopeR design includes a variety of features described below. We intend to continue to make IsotopeR's graphical user interface (GUI) simple and intuitive, and we welcome any feedback that will help improve it.
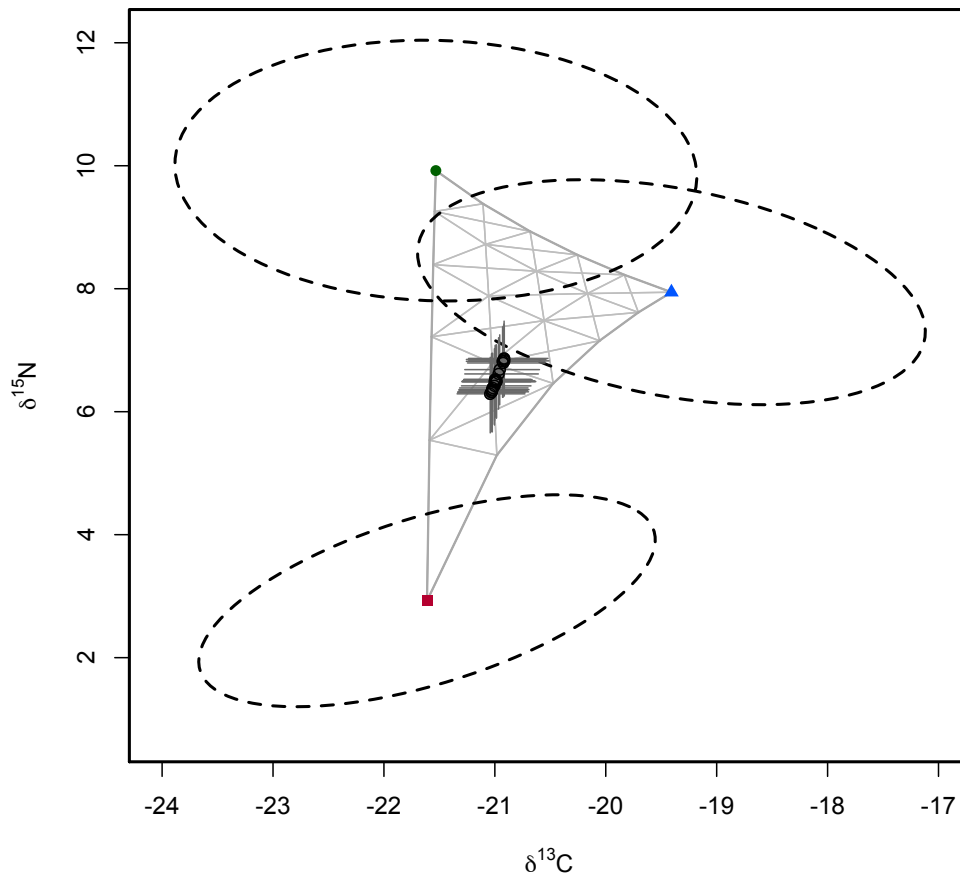
2　Installing IsotopeR

- Install JAGS for Mac http://sourceforge.net/projects/mcmc-jags/ or for Windows http://sourceforge.net/projects/mcmc-jags/files/JAGS/4.x/Windows/.

- Mac users with R before v3.0 must install tcltk software http://cran.r-project.org/bin/macosx/tools/ and all mac users must also install XQuartz 2.7.8 http://www.xquartz.org/. For smoother operation, update your computer to R 3.2.2 for Mac OS X 10.9 or higher.

- Install "IsotopeR" and the following dependencies (packages: *fgui, runjags, ellipse, plotrix, colorspace*) using the following code in R:

  ```
  > install.packages("IsotopeR", dep=T)
  ```

3　Formatting input data files

When formatting your input data files (names and contents of each file described below) use the same format as the example data. The easiest way to ensure the data is entered correctly is to cut and paste your raw data into each .csv file (not .xls) provided here and save.

3.1　Mixtures

The first n columns in this input data file are the isotope values for consumers' tissues, where n is the number of stable isotope systems used in the analysis. The last two columns designate the group (e.g., sub-population, sex) and individual assignments. If there is no group structure, then column $n+1$ will contain "1" for all individuals (see example below). If designating multiple groups, the group identity will be determined by the variable in the column. Individuals in the second group should be designated as "2", the third group as "3", etc. Individuals in second should be entered after individuals in the first group and so on. The last column identifies each individual. If you have no repeated measurements for individuals, then each individual should be designated by a unique integer (e.g., 1, 2, 3,...; see example below), whereas individuals with repeated measures should be designated using the same number (e.g., 1, 1, 1, 2, 2, 2,...). Note that IsotopeR does not require consecutive integers for the individual column; however the estimated values are reported as consecutive, so it may be difficult to match the input to the output if you use nonconsecutive integer labels in the data file.

d13C	d15N	group	individual
-20.7	6.2	1	1
-21.2	6.5	1	2
-21	6	1	3
-21	6.7	1	4
...	...	...	...

3.2   Sources

Each source is a sample of foods consumers are known to eat, which could be a sample of the same species or an aggregate of similar taxa. Although a number of studies have put forward strategies for aggregating foods (and we concur that justification is necessary), such grouping is dependent on the question of interest. The first n columns in this input data file are the isotope values for each sampled food corrected for isotopic discrimination (i.e., add the appropriate discrimination factor to each observation), where n is the number of isotopes used in the analysis. It is important to make sure that isotope values are in the same order as the mixture data file (e.g., column 1 in both the Mixtures and Sources file should contain $\delta^{13}C$ values). The next column $(n+1)$ identifies the source to which the sampled food items belong. Unlike other columns, you can use the actual names (text, not integers) of the food sources (e.g., Source 1, Source 2, Source 3). The last column (subsource) identifies different foods within each source; this feature assigns equal weight to each specified subsource.

d13C	d15N	source	subsource
-21.7	2.9	Source 1	1
-22	2.6	Source 1	1
-21.6	3	Source 1	1
-21.8	2.9	Source 1	1
-21.4	3	Source 1	1
-21.5	10	Source 2	1
-21.6	9.8	Source 2	1
...	...	...	...

3.3   Concentrations

The first n columns in this input data file are the digestible elemental concentrations for each sample, where n is the number of digestible elemental concentrations used in the analysis (e.g., [C], [N]). Columns with elemental concentrations need to match those in the Sources file (e.g., column 1 in this file and Sources files contain [C] and $\delta^{13}C$ values, respectively). Column $n+1$ identifies the source (use the same text as in the Mixtures, Sources, Concentrations files) in which the set of concentrations belong. The last column links sampled food concentrations to each subsource. This feature assigns equal weight to each subsource's digestible elemental concentrations and should be consistent with the Sources file.

3.4   Discrimination Error

This input data file contains the standard deviations associated with the estimated average discrimination factors from diet studies. The first n columns are the standard deviations associated

[C]	[N]	source	subsource
25	5	Source 1	1
26	4	Source 1	1
25	4	Source 1	1
24	3	Source 1	1
25	5	Source 1	1
40	10	Source 2	1
40	10	Source 2	1
...	...	...	...

with each mean discrimination value for the associated stable isotope ratio. The last column denotes the source identification for the standard deviations, which must match those specified in the Mixtures and Sources files.

delta13C	delta15N	source
1	0.5	Source 1
1	0.5	Source 2
1	0.5	Source 3

3.5 Measurement Error

This is the measurement error associated with mass spectrometry. This input data file contains all isotopic measurements for standards. Isotope values need to be in the same order as other data files (e.g., column 1 in Mixtures, Sources, and Measurement Error files contain $\delta^{13}C$ values).

d13C	d15N
-12.7	5.5
-12.6	5.5
-12.5	5.5
-12.5	5.4
-12.6	5.5

4 Running IsotopeR

4.1 Opening the IsotopeR GUI

After installing the packages above for the first time in R, you can open the IsotopeR GUI with the following code (comments about the command follow the "#" symbol):

```
> library(IsotopeR) # loads packages (Fig. 1)
> IsotopeR() #open Graphical User Interface (Fig. 2)
```

If correctly installed, the main menu for the IsotopeR GUI will appear as shown in Figure 1.
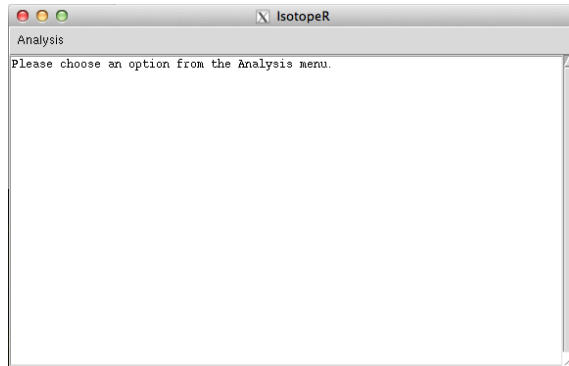
Figure 1: Main window for the IsotopeR GUI

4.2 Loading input data files

To load input data files, select **Analysis → New Run** from the menu (Figure 1). A data input screen will appear (Figure 2). Here, you will enter your stable isotope data.
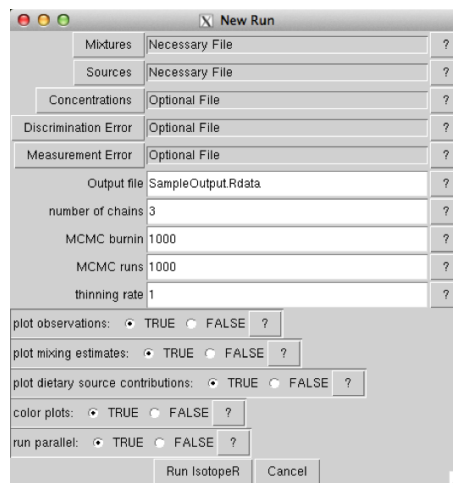

Figure 2: Data input screen for the IsotopeR GUI

To enter the input data files, click on each button on the left hand side of the IsotopeR GUI (Figure 2); clicking each button will navigate you to your hard drive and allow you to select the appropriate input data file (Figure 3).
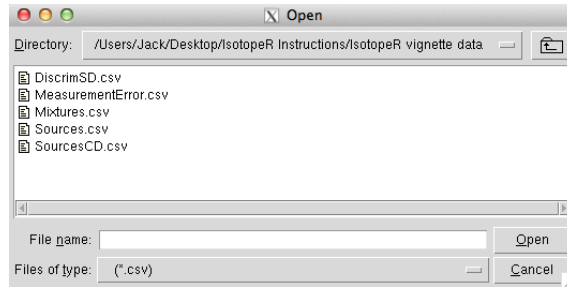
Figure 3: Navigation window for selecting input data files

4.3 Control parameters and plotting options

Options to adjust control parameters and plots are located below the file load section of the IsotopeR GUI (Figure 2). Each option can be turned on (TRUE) and off (FALSE).

Number of [Markov] chains: The number of independent Markov chains for sampling the parameter space and computing posterior quantities.

MCMC burn-in: The length of the chain discarded at the beginning of the run, so that the effect of initial values on the posterior inference is minimized. This is interpreted as the length of time it takes for the MCMC to stabilize.

MCMC runs: The total number of iterations per chain, excluding burn-in.

Thinning rate: This feature reduces the sample size of each MCMC run by keeping every kth simulated draw from each chain. When thinning equals 1, the entire chain is saved.

Plot observations: Plots all raw isotope values for consumers and food items.


7
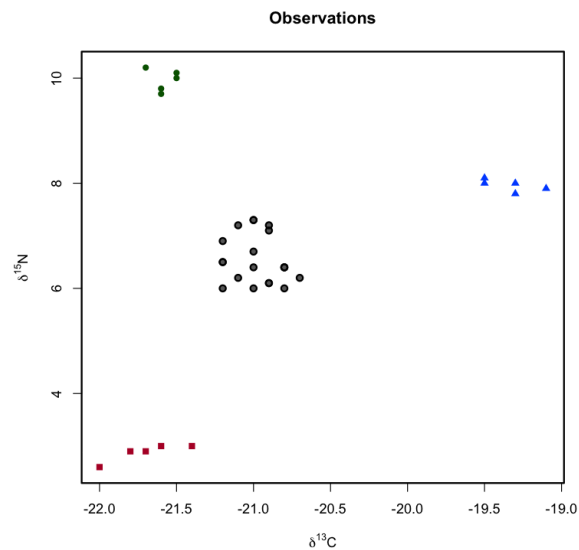
Plot mixing estimates: Plots the estimated isotopic mixing space (the area contained in the space formed by lines connecting sources in a multivariate plot). The isotopic mixing space depicts the mechanistic relationship between consumers and their foods (i.e., the environment of consumers). Sources are denoted by dashed ovals (2 SD) and gray lines loosely define how proportional diet contributions were calculated. Note that mixtures are displayed differently in the observation and estimated mixing space plots because IsotopeR estimated these isotope values for consumers (with error bars denoting the 95% credible interval of the estimate) as well as their sources, as they are random variables, not fixed.

Estimated Mixing Space
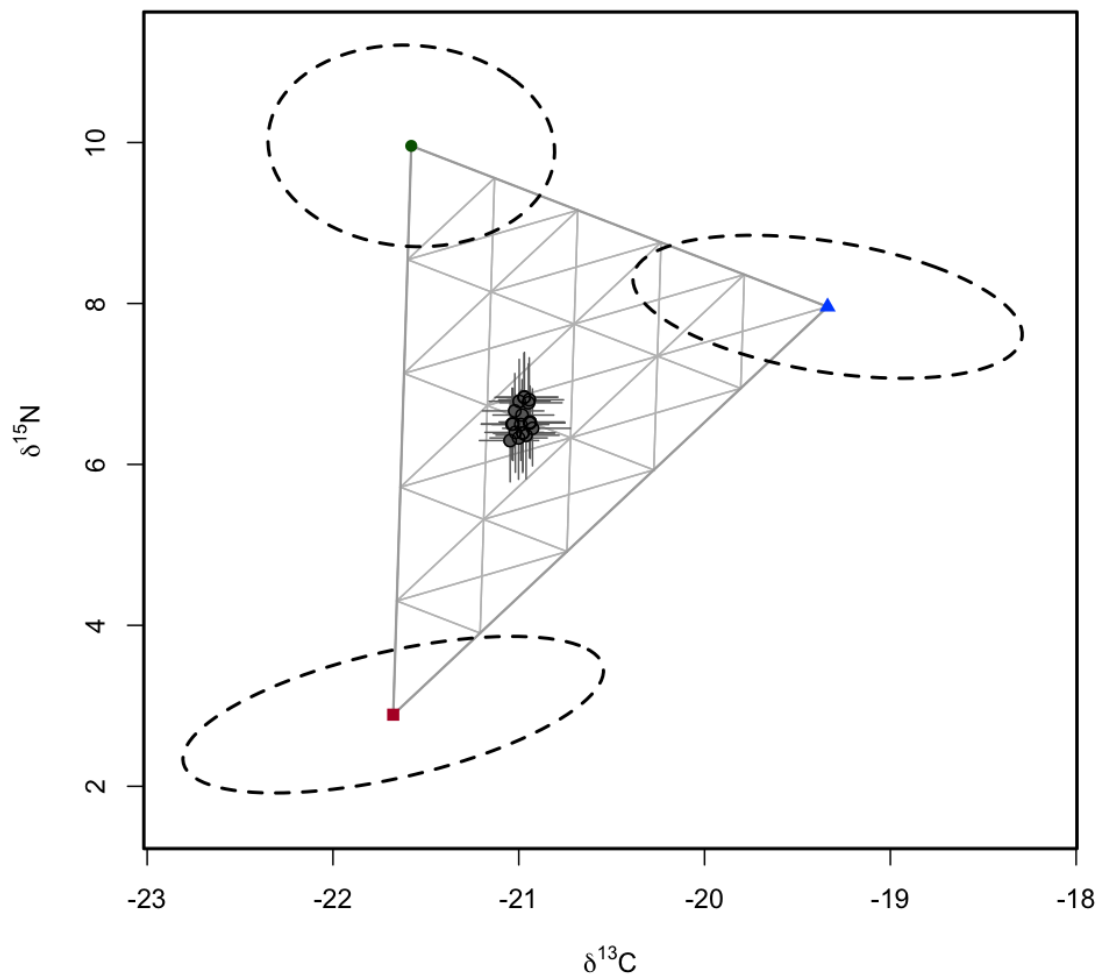
Plot dietary source contributions: Plots the mean of the median posterior probability distribution for each major food source. Generally, steep and narrow posteriors for populations or groups denote low variability of resource use, whereas short and wide posteriors suggest higher variability of resource use. A large range of median posteriors for individuals, with little overlap, suggests consumers have heterogeneous diets. The width of posteriors is also related to the variability of isotope values for each source and their positions relative to one another in the isotopic mixing space. An increase in the isotopic variation of a source or the overlap of isotope values for sources may increase collinearity of sources, decreasing the precision of dietary estimates, increasing credible intervals. Such overlap can lead to the multimodality observed in posterior distributions.
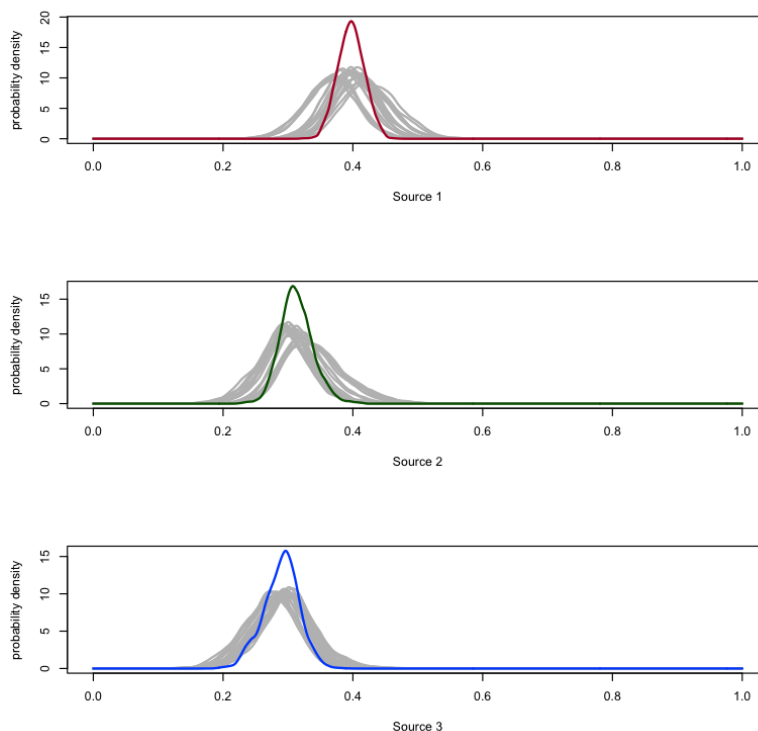
Run parallel: This boolean flag determines whether the MCMC chains should be run on separate cores of the processor (parallel=TRUE) or not (parallel=FALSE). Running parallel chains requires that the number of MCMC chains is greater than 1 and that the machine running the MCMC has multiple cores. Note that the Deviance Information Criterion (DIC) will not be reported when chains are run in parallel.

4.4 Parameter estimates

Parameter estimates will be automatically written to a text file (SampleOutput.txt) and an image file (SampleOutput.Rdata file), both located in the current working directory in R unless another output directory is defined by the user in the GUI (Figure 2). The .Rdata file name is specified in the field **Output file** (in the middle of the IsotopeR GUI; Figure 2). The files are formatted in a matrix with rows given by the parameter names (defined below). The first two columns are the mean and standard deviation of the median posterior probability distribution. Credible intervals described as quantiles (2.5%, 25%, 50%, 75%, 97.5%) for this sampling distribution are reported in respective columns, followed by the Rhat values, which is a metric of convergence that should be less than 1.2 or the model should be rerun with a longer MCMC chain or more MCMC runs. The Rhat statistic assumes that the posterior distributions are normally distributed. When this assumption is violated, as often occurs when dietary proportions are near the boundary of their [0,1] support, these values may not be reliable indicators of convergence. If a dietary estimate is close to zero, then the food is not an important dietary source. We suggest that you remove the isotope values (and digestible concentrations if applicable) for this food from the input data files and rerun IsotopeR. Before removing the source however, we suggest users confirm the data was entered correctly (e.g., be sure it was corrected for discrimination). We note that the Deviance Information Criterion (DIC) can be used to conduct Bayesian model selection (at the end of output file). For instance, a user can compare model fits using isotope values for other sources or including random effects.

Parameters:

mu.source(z,i): Mean isotope value for food source z, isotope i

sd.source(z,i): Standard deviation of the isotope value for food source z, isotope i

rho.mat(z,i,j): Correlation between isotopes i and j for food source z

mu.conc(z,i): Mean elemental concentration for food source z, isotope i

sd.conc(z,i): Mean of the standard deviation for food source z, isotope i

mu.mix(x,i): Isotope value for individual x, isotope i

p(x,z): Proportional dietary contribution for individual x, food source z

p.pop(z): Population-level proportional dietary contribution for food source z

sd.res(i): The residual error term (standard deviation) for isotope i

sd.me(i): Measurement error (standard deviations) for isotope i

p.group (k,z): Proportional dietary contribution for group k, food source z

deviance: Twice the negative log-likelihood of the model; a measure of the model's fit to the data, where the lower the measure, the better the model's fit

4.5 Error messages

Errors are often associated with JAGShttp://sourceforge.net/projects/mcmc-jags/. Generally, these errors are often related to the model not converging. If you receive an error, rerun your analysis with more MCMC chains or use a higher thinning rate. A gentle introduction to JAGS is provided in the document by N. Thompson Hobbes: An Ecological Modeler's Primer on JAGS. JAGS model syntax is compatible with the BUGS language. Users unfamiliar with the BUGS language can find many tutorials at the WinBUGS site: http://www.mrc-bsu.cam.ac.uk/bugs/.

5 Examples

IsotopeR can estimate dietary parameters for ≥ 2 isotope ratios (n) and $\geq n+1$ sources. Estimating proportional dietary contributions for $> n+1$ sources generally does not result in unique solutions (i.e., they are underdetermined mixing models). We restrict the following examples to 2- and 3-isotope models with 3 sources. Please download the input data files for the following analysis available at input data files.

5.1 2-isotope model

5.1.1 Standard model
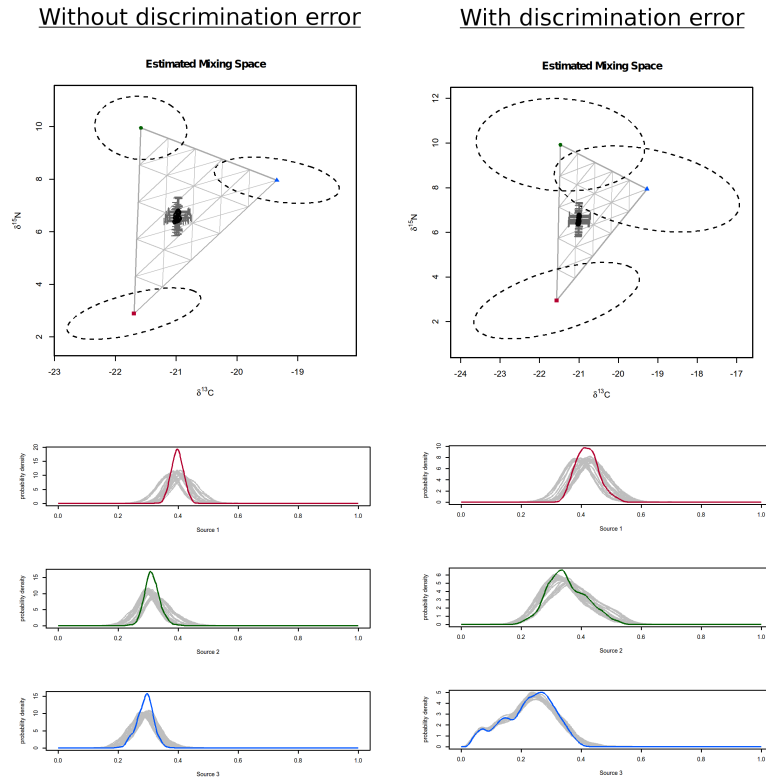
- Load the following input data files from the 2-isotope folder: Mixtures.csv, and Sources.csv.

- Keep IsotopeR in default mode by not changing any entries for control parameters or the plots (all should read TRUE).

- Click the **Run IsotopeR** box.

- The estimation procedure will begin as indicated by the text: "Calling the simulation... (this might take some time)". When the estimation process has finished, plots will appear (pg. 7-9), and parameter estimates and diagnostic output and parameter estimates will be provided in the R console (pg. 10).

5.1.2 Discrimination error

Stable isotope discrimination factors (small offsets of stable isotope values between dietary sources and animal tissues due to metabolic and digestive processes, as determined from studies of animals on controlled diets; e.g., expressed as Δ^{13}C, Δ^{15}N, Δ^{34}S) **NEED to be added to the isotope values of each food source before running each analysis**. Including the variation associated with discrimination factors (Discrimination Error) derived from controlled studies is also important to accurately estimate proportional dietary contributions for consumers using IsotopeR.

- Load the following input data files from the 2-isotope folder: Mixtures.csv, Sources.csv, and DiscriminationError.csv.

- Keep IsotopeR in default mode by not changing any entries for the plots or control parameters (all should be TRUE) and click the **Run IsotopeR** box.

- Examine the output and plots. The estimated sources in the isotopic mixing space have become larger (top right) and posteriors (bottom right) are wider than an analysis without discrimination error (left plots).

5.1.3 Measurement error

Although the effect of measurement error on parameter estimates can often seem negligible, this source of isotopic variation is inherent to stable isotope analysis. To include this source of error in your analysis, follow the directions under Formatting input data files (Section 3).
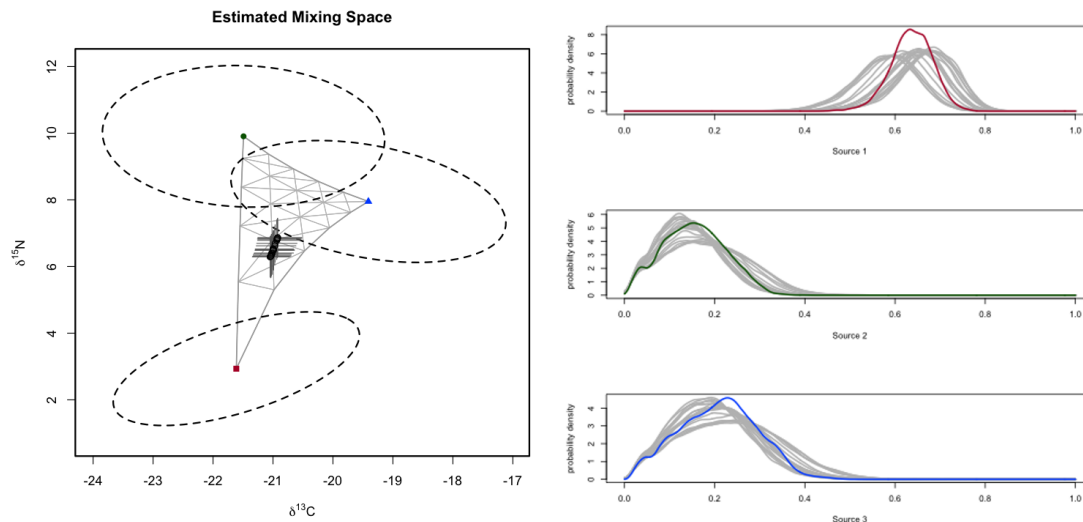
- Load the following input data files from the 2-isotope folder: Mixtures.csv, Sources.csv, DiscriminationError.csv, and MeasurementError.csv

- Keep IsotopeR in default mode and click the **Run IsotopeR** box

- Examine the output and plots. The plots and population estimates look very similar to the last analysis except that there are error bars on mixtures in the Observations plot. Error

bars appear on raw mixtures because including measurement error specifies error associated with each mixture prior to the analysis.

5.1.4 Digestable elemental concentrations

Not accounting for differences in the digestibility of elements among food sources (concentration dependence) can severely bias dietary estimates. In particular, the relative differences in the elemental concentrations between sources of plant and animal matter digested by animals can cause the mixing lines to bend between sources, creating a nonlinear mixing space, which effects the relationship (and parametrization) between consumers and their foods. Here, we add new digestible concentration data to our analysis.
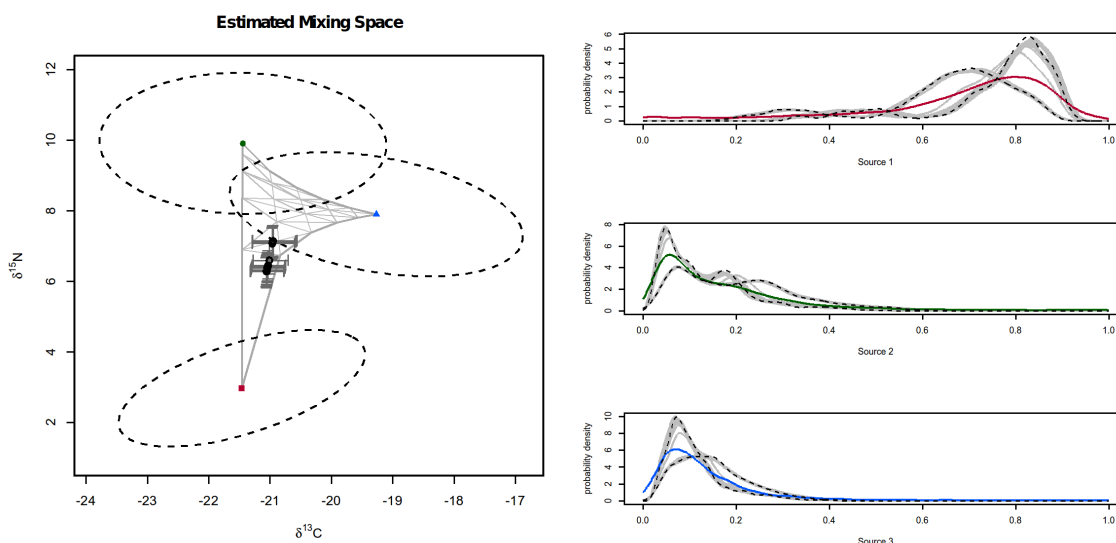
- Load the following input data files from the 2-isotope folder: Mixtures.csv, Sources.csv, DiscriminationError.csv, MeasurementError.csv, and SourcesCD.csv.

- Change MCMC runs from 1000 to 5000 and click the **Run IsotopeR** box.

- Examine the output and plots. Notice that the estimated isotopic mixing space has morphed dramatically as a result of including the concentration dependence data in the analysis. Changes to the isotopic mixing space more accurately illustrates the relationship between consumers and their foods (their environment), leading to more accurate parameter estimates as shown in the posteriors plot below.

5.1.5 Group assignments

IsotopeR has the ability to also specify groups (e.g., sexes) as random effects.

- Load the following input data files from the 2-isotope folder: Mixtures_groups.csv, Sources.csv, Concentrations.csv, DiscriminationError.csv, MeasurementError.csv

- Change MCMC runs from 1000 to 5000 and click the **Run IsotopeR** box

- Examine the output and plots. The isotopic mixing space (left) is similar to the previous mixing space with no specified groups, except that two groups are evident in this plot. Posteriors for each group (dashed lines) are included in the posteriors plot (right) and estimates for these posteriors follow p.group (see section 4.4) in the output.


5.2 3-isotope model

IsotopeR also has the ability to graphically display both raw isotope values and estimated mixtures and sources in rotational 3-D plots. For the following analysis, we use the same data from the 2-isotope, 3-source analysis but include measurements of $\delta^{34}S$ for mixtures and sources, digestible elemental concentrations for S, and discrimination error. Each of the following 3-D plots are rotational when viewed in R but saved as snapshots for display purposes. These 3-D plots are made using the "rgl" package. To install:

```
> install.packages("rgl", dep=T)
```

- Load the following input data files from the 3-isotope folder: Mixtures.csv and Sources.csv.

- Keep IsotopeR in default mode by not changing any entries for the plots or control parameters (all should be TRUE).

- Click the **Run IsotopeR** box.

- Examine the output and plots. Notice that posteriors and estimates are similar to the standard 2-isotope model. Unlike the 2-isotope model, the Observations plot (below) can be rotated, illustrating the raw stable isotope data for consumers and their major foods in a 3D isotopic mixing space.
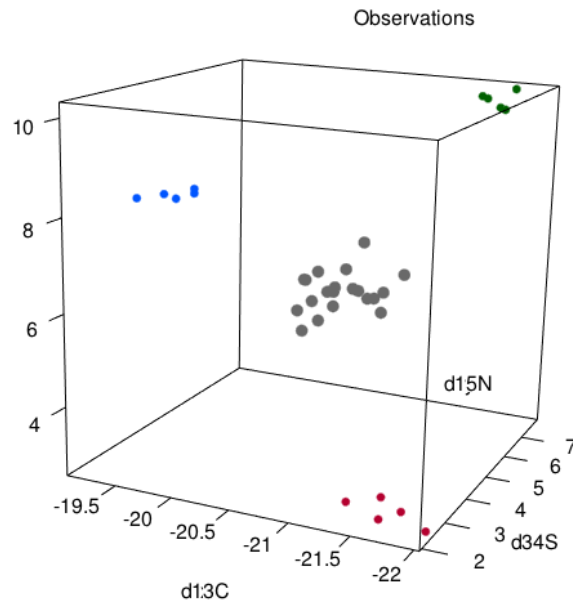
- Similar to the isotopic mixing space for the 2-source model, this plot illustrates the relationship between consumers and their major foods. This plot rotates but currently does not include the gray lines showing how dietary parameters were estimated.

Estimated Mixing Space
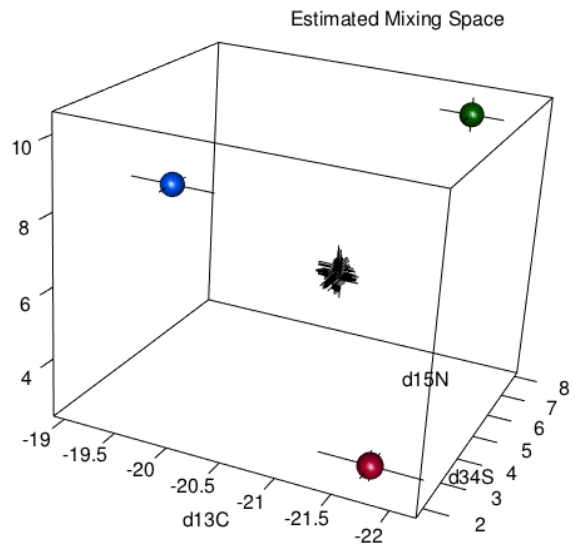
- Now, load the following input data files from the 3-isotope folder to estimate groups, including digestible [S], and Δ^{34}S error: Mixtures_groups.csv, Sources.csv, Concentrations.csv, Measurement Error.csv, and Discrimination Error.csv

- Change MCMC runs from 1000 to 5000 and click the **Run IsotopeR** box

- Examine the output and plots. Mixtures and proportional dietary contributions were estimated in two groups (see below).

Estimated Mixing Space
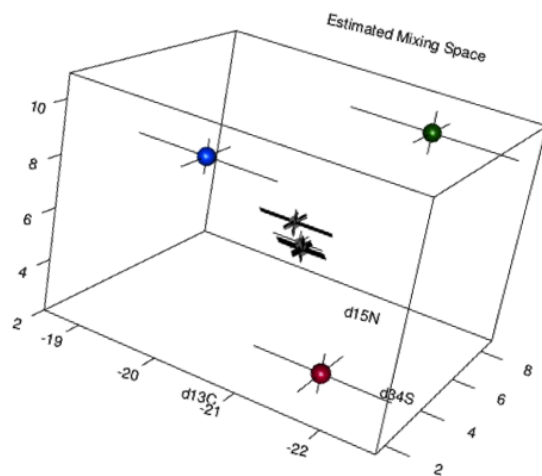
- Although this mixing model was overdetermined (3 isotopes & 3-sources), most posteriors are narrower than the 2-isotope 3-source model, indicating more precise estimates than parameters previously estimated (pg. 14).