

# Basic Analysis of Simplex output

*Ken Locey*

*October 2015*

## OVERVIEW

This R Markdown document is designed to be opened and ran in the RStudio. The chunks of code below allow for basic analysis of **simplex** output. This includes univariate and multivariate relationships, and graphical exploration.

For an R-based introduction to multivariate analysis: <https://little-book-of-r-for-multivariate-analysis.readthedocs.org/en/latest/src/multivariateanalysis.html>

## SETUP

### A. Clear and set the working directory

```
rm(list=ls())
getwd()
setwd("~/GitHub/simplex")
```

### B. Import packages; install if needed

```
#install.packages("vegan") # Example of how an install can be done
require("vegan")
require("car")
```

### C. Import simulated data generated by simplex models

As it is iterating through randomly assembled models, simplex writes its output to six .csv files. For each file, each row corresponds to a single model. Consequently, the *i*th row in each file corresponds to *i*th model that was assembled and run by simplex. Let's import the simulated data files.

```
# A table where each columns corresponds to a state variable or model output.
simplex.dat <- read.csv("~/GitHub/simplex/results/simulated_data/examples/SimData.csv")

# Replacing 0's with 1's to allow log-transforms below
simplex.dat <- simplex.dat[complete.cases(simplex.dat), ]

simplex.dat$total.abundance[simplex.dat$total.abundance<=0] <- 0.00001
simplex.dat$N.max[simplex.dat$N.max<=0] <- 0.00001
simplex.dat$species.richness[simplex.dat$species.richness<=0] <- 0.00001
simplex.dat$resource.particles[simplex.dat$resource.particles<=0] <- 0.00001
simplex.dat$resource.concentration[simplex.dat$resource.concentration<=0] <- 0.00001
simplex.dat$Whittakers.turnover[simplex.dat$Whittakers.turnover<=0] <- 0.00001
simplex.dat$Jaccards.dissimilarity[simplex.dat$Jaccards.dissimilarity<=0] <- 0.00001
```

```
simplex.dat$total.abundance <- log(simplex.dat$total.abundance)
simplex.dat$N.max <- log(simplex.dat$N.max)
simplex.dat$Whittakers.turnover <- log(simplex.dat$Whittakers.turnover)
simplex.dat$Jaccards.dissimilarity <- log(simplex.dat$Jaccards.dissimilarity)
simplex.dat$species.richness <- log(simplex.dat$species.richness)
simplex.dat$resource.particles <- log(simplex.dat$resource.particles)
simplex.dat$resource.concentration <- log(simplex.dat$resource.concentration)
```

## UNIVARIATE ANALYSES

Perhaps we want to ask whether **simplex** produces (i) relationships that are well-known to occur in ecological systems (ii) auto-correlated relationships between independent variables (iii) or novel relationships of use to a specific study or question. Each of these questions has its particular use, but rather than generate one x-y relationship after another, we generate an entire field of relationships. That is, given x variables, we can explore  $x*(x-1)$  x-y relationships.

Let's begin by examining relationships between some physical variables. As we'll see, the only physical variables that appear to be correlated are ecosystem residence time and the rate of flow.

```
# Physical and metacommunity variables
png(filename="~/GitHub/simplex/results/figures/examples/PhysicalUnivariate.png")

phys.dat <- as.matrix(subset(simplex.dat,
                           select = c(flow.rate,
                                       barriers,
                                       width,
                                       height,
                                       amplitude)))

scatterplotMatrix(phys.dat, main="Physical data", diagonal = "none")
dev.off()
```

Next, let's explore relationships between some resource-related variables. As we'll see, the number of resource particles, their concentration, resource richness or the number of resource types, and resource diversity (a combination of resource richness and the variance in abundance among resources) are all highly and positively correlated.

```
# Resource variables
png(filename="~/GitHub/simplex/results/figures/examples/ResourceUnivariate.png")

res.dat <- as.matrix(subset(simplex.dat,
                           select = c(resource.concentration,
                                       shannons.resource.diversity,
                                       resource.richness,
                                       resource.particles)))

scatterplotMatrix(res.dat, main="Resource data", diagonal = "none")
dev.off()
```

We can also explore relationships between physiological variables, noticing that, biomass production in Carbon, Nitrogen, and Phosphorus are all positively and strongly correlated. This is largely because **simplex** (as of 26 October 2015) does not include any explicit stoichiometry. Also notice that an interesting life

history trade-off arises in **simplex** however, that is, increasing growth rate leads to increasing metabolic maintenance.

```
# Physiological variables
png(filename="~/GitHub/simplex/results/figures/examples/PhysioUnivariate.png")

physio.dat <- as.matrix(subset(simplex.dat,
                              select = c(biomass.prod.N,
                                           avg.per.capita.N.efficiency,
                                           max.growth.rate,
                                           max.met.maint)))

scatterplotMatrix(physio.dat, main="Physiological data", diagonal = "none")
dev.off()
```

Finally, let's explore relationships between some diversity related variables. As we can see different evenness measure basically reflect each other, as do different measures of species turnover. Likewise, we see a strong positive relationships between total abundance and species richness and between total abundance and the abundance of the most abundant species; which we should expect.

```
# Diversity-related variables
png(filename="~/GitHub/simplex/results/figures/examples/DiversityUnivariate.png")

diversity.dat <- as.matrix(subset(simplex.dat,
                                  select = c(total.abundance,
                                              N.max,
                                              species.richness,
                                              simpson.e,
                                              Whittakers.turnover,
                                              Jaccards.dissimilarity)))

scatterplotMatrix(diversity.dat, main="Diversity data", diagonal = "none")
dev.off()
```

## Summary statistics

We can use the 'sapply' function to generate summary statistics (mean, variance, etc.) column by column for any of our **simplex** output. For example, analyzing phys.dat as a data.frame:

```
sapply(as.data.frame(phys.dat), mean) # sample mean
```

```
## flow.rate barriers width height amplitude
## 0.5113316 2.0417755 51.6710183 27.1671018 0.2412533
```

```
sapply(as.data.frame(phys.dat), var) # sample variance
```

```
## flow.rate barriers width height amplitude
## 9.586341e-02 1.809769e+00 2.515001e+03 6.342966e+02 2.369947e-02
```

```
sapply(as.data.frame(phys.dat), sd) # sample standard deviation
```

```
## flow.rate barriers width height amplitude
## 0.3096182 1.3452764 50.1497900 25.1852459 0.1539463
```

```
sapply(as.data.frame(phys.dat), median) # sample median
```

```
## flow.rate barriers width height amplitude
## 0.5 2.0 40.0 20.0 0.2
```

## HIGHLY CORRELATED VARIABLES

Suppose we want to find the most highly correlated variables in our data. First, let's define a function to return the x most highly correlated variables in a dataframe.

Then, let's call our function to find the x most highly correlated variables.

```
dat <- cbind(phys.dat, physio.dat, res.dat, diversity.dat)
mosthighlycorrelated(dat, 15)
```

	First.Variable	Second.Variable	Correlation
## 360	Whittakers.turnover	Jaccards.dissimilarity	0.9973543
## 220	shannons.resource.diversity	resource.richness	0.9606565
## 320	species.richness	simpson.e	-0.8396928
## 319	N.max	simpson.e	-0.7762931
## 318	total.abundance	simpson.e	-0.7608577
## 299	total.abundance	species.richness	0.7151727
## 280	total.abundance	N.max	0.6803611
## 160	max.growth.rate	max.met.maint	0.5656628
## 174	width	resource.concentration	-0.5641684
## 267	flow.rate	N.max	-0.4808350
## 300	N.max	species.richness	0.4618625
## 232	height	resource.particles	0.4443530
## 358	species.richness	Jaccards.dissimilarity	0.4183885
## 230	barriers	resource.particles	-0.4170529
## 339	species.richness	Whittakers.turnover	0.3734704

## Variance partitioning.

A common question to ask is whether physical variables (geographic distance, area, volume, flow rate, etc.) has a larger influence on ecological diversity than, say, resource-related variables.

One way to address this question is to use variance partitioning. A note of caution, typing 'help(varpart)' reveals that R only uses (simple) linear regression when there is one response variable and uses redundancy analysis ordination (RDA) for two or more response variables.

Executing the code below shows that, in **simplex** models, the phys variables explain more variation in total abundance, and species richness, evenness, and turnover than do resource related variables.

