Abundance analyses

Jeff Oliver

August 23, 2018

Testing for effects of sampling method on abundances, as well as how size influences abundance

Methods

Setup

Load data and third-party packages.

```
library("tidyr")    # for gather
library("ggplot2") # for plotting
library("nlme")     # for mixed-effect models (lme)
bioscan <- read.csv(file = "data/BioScanData.csv")
species.data <- read.csv(file = "data/species-data.csv")

# Drop any rows missing data
bioscan <- na.omit(bioscan)
```

Data wrangling

Only want to use data for those sites where we have information for both collection types.

```
# Identify sites with data for each of the two collection methods
pollard.sites <- bioscan$Site.Number[bioscan$Collection.Method == "Pollard Walk"]
malaise.sites <- bioscan$Site.Number[bioscan$Collection.Method == "Malaise"]

# Identify sites with data for *both* collection methods
sites.with.both <- intersect(x = pollard.sites, y = malaise.sites)

# Reduce dataset to only those sites with both types of data
bioscan <- bioscan[bioscan$Site.Number %in% sites.with.both, ]
rownames(bioscan) <- NULL
```

We also need to transform our data so (1) each row corresponds to a single species and (2) the species size data is also included.

```
# Transform data to long format
bioscan.long <- bioscan %>%
  gather(key = "Species", value = "Abundance", -c(1:4))

# Add species data
bioscan.long <- merge(x = bioscan.long,
                      y = species.data,
                      by.x = "Species",
                      by.y = "species")
# Use the mean of the minimum and maximum as single value for size
bioscan.long$size <- apply(X = bioscan.long[, c("opler.wright.min", "opler.wright.max")],
```

```
                              MARGIN = 1,
                              FUN = function(x) {mean(x)})
```

Statistical analysis

Effect of sampling type

Test the effect of collection method on abundance, including site as a random intercept effect.

$$Abundance = \beta_0 + \beta_1 \times CollectionMethod + b_{0i}$$

where b_{0i} is a random intercept for the i^{th} site.

```
# Run simple model, testing the null hypothesis of intercept = 0 (no difference
# in abundance)
simple.model <- lme(Abundance ~ Collection.Method,
                    random = ~1|Site.Number,
                    data = bioscan.long)
simple.tTable <- summary(simple.model)$tTable
```

Effect of species size

Test a model where size, as well as the interaction between collection method and size, are included in the model to predict abundance.

$$Abundance = \beta_0 + \beta_1 \times CollectionMethod + \beta_2 \times Size + \beta_3 \times CollectionMethod \times Size + b_{0i}$$

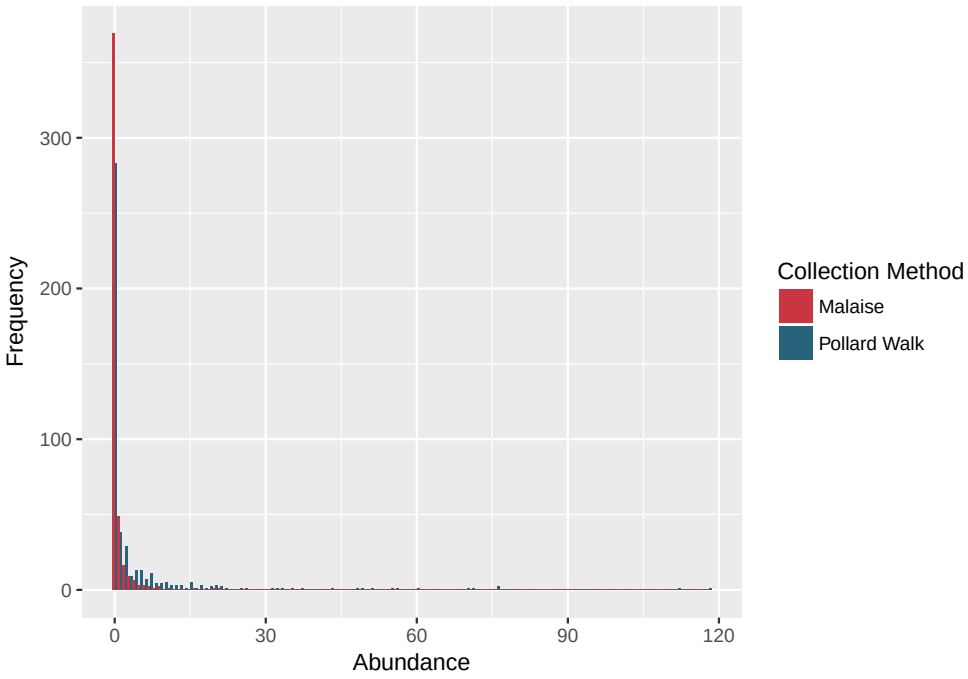
where b_{0i} is a random intercept for the i^{th} site.

```
# Run a more complex model, with size as a fixed effect, as well as the
# interaction between size and Collection.Method
size.model <- lme(Abundance ~ Collection.Method + size + Collection.Method*size,
                  random = ~1|Site.Number,
                  data = bioscan.long)
size.tTable <- summary(size.model)$tTable
```

Results

On average, abundance was higher in Pollard walk surveys than in Malaise traps ($t = 6.352$, $p < 0.001$). I'm not sure this plot is the best way to visualize this...

```
abundance.distribution <- ggplot(data = bioscan.long,
                                 mapping = aes(x = Abundance, fill = Collection.Method)) +
  geom_histogram(binwidth = 1, position = "dodge") +
  xlab(label = "Abundance") +
  ylab(label = "Frequency") +
  scale_fill_manual(values = c("#CA3542", "#27647B"), name = "Collection Method")
print(abundance.distribution)
```



Neither species size nor the interaction between size and collection method have a significant effect on abundance:

Coefficient	Estimate	t	p
Intercept	0.986	1.009	0.313
Collection method	4.276	3.21	0.001
Size	-0.008	-0.472	0.637
Collection method x Size	-0.01	-0.42	0.675