

Principal Components

J.P. Meagher

6 October 2017

In order to get some ‘quick and dirty’ ancestral reconstruction results, the spectrogram time registration step will be skipped for the time being. Previous analysis has suggested that this will not have much impact on the results.

Packages

```
library(batwork)
library(tidyverse)
library(magrittr)
library(RColorBrewer)
```

Data

```
observed_frequencies <- get_spectrogram_details(
  mexican_bat_calls %>% use_series('call') %>% extract2(1),
  detail = 'freq')

band <- observed_frequencies > 9000 & observed_frequencies < 212000

df <- mexican_bat_calls %>%
  mutate(time = map(calls, get_spectrogram_details, detail = 'time'),
    psd = map(calls, get_psd)) %>%
  mutate(restricted = map(psd, extract, , band)) %>%
  mutate(restricted = map(restricted, log10)) %>%
  mutate(restricted = map(restricted, multiply_by, 10)) %>%
  mutate(smoothed = map(restricted, waveslim::denoise.dwt.2d, J = 2)) %>%
  mutate(regularised = map(smoothed, grid_interpolation)) %>%
  select(bat, species, sex, regularised)

f <- seq(observed_frequencies %>% extract(band) %>% min,
  observed_frequencies %>% extract(band) %>% max,
  length.out = df %>% use_series('regularised') %>% extract2(1) %>% ncol)

t <- seq(0, 1,
  length.out = df %>% use_series('regularised') %>% extract2(1) %>% nrow)
```

Resampling

When obtaining components for spectral density curves a resampling procedure was used to obtain stable, significant components. Something similar will be attempted for spectrograms in the future. For the time being, the steps required to obtain one of these samples are laid out, but the resampling not implemented.

```
draw <- df %>%
  group_by(bat) %>%
  sample_n(1) %>%
  ungroup %>%
  group_by(species) %>%
  sample_n(4) %>%
  ungroup
```

For this preliminary run of the model it is not necessary to implement a full resampling procedure. Thus further analysis will be done on the whole sample.

Mean Spectrogram

Calculate the mean spectrogram of bat calls from this sample.

```
mean_spectrogram <- df %>%
  select(regularised) %>%
  unnest %>%
  use_series('regularised') %>%
  array(dim = c(50, 100, df %>% nrow)) %>%
  apply(c(1,2), mean)
```

```
image(t, f %>% divide_by_int(1000),
  mean_spectrogram,
  main = 'Mean Spectrogram',
  xlab = 'Time', ylab = 'Frequency (kHz)',
  col = brewer.pal(9, 'Spectral'))
```

Principal Components Analysis

By vectorising the spectrograms a principal components analysis can be performed.

```
pca <- df %>%
  select(regularised) %>%
  unnest %>%
  use_series('regularised') %>%
  array(dim = c(50*100, df %>% nrow)) %>%
  t %>%
  prcomp
```

```
variation <- pca %>% use_series(sdev) %>% raise_to_power(2)
```

```
components <- pca %>% use_series(rotation) %>% extract(,1:10) %>% c %>% array(dim = c(50, 100, 10))
```

```
image(t, f %>% divide_by_int(1000),
  components %>% extract(,1),
  main = 'Example Component',
  xlab = 'Time', ylab = 'Frequency (kHz)',
  col = brewer.pal(9, 'Spectral'))
```

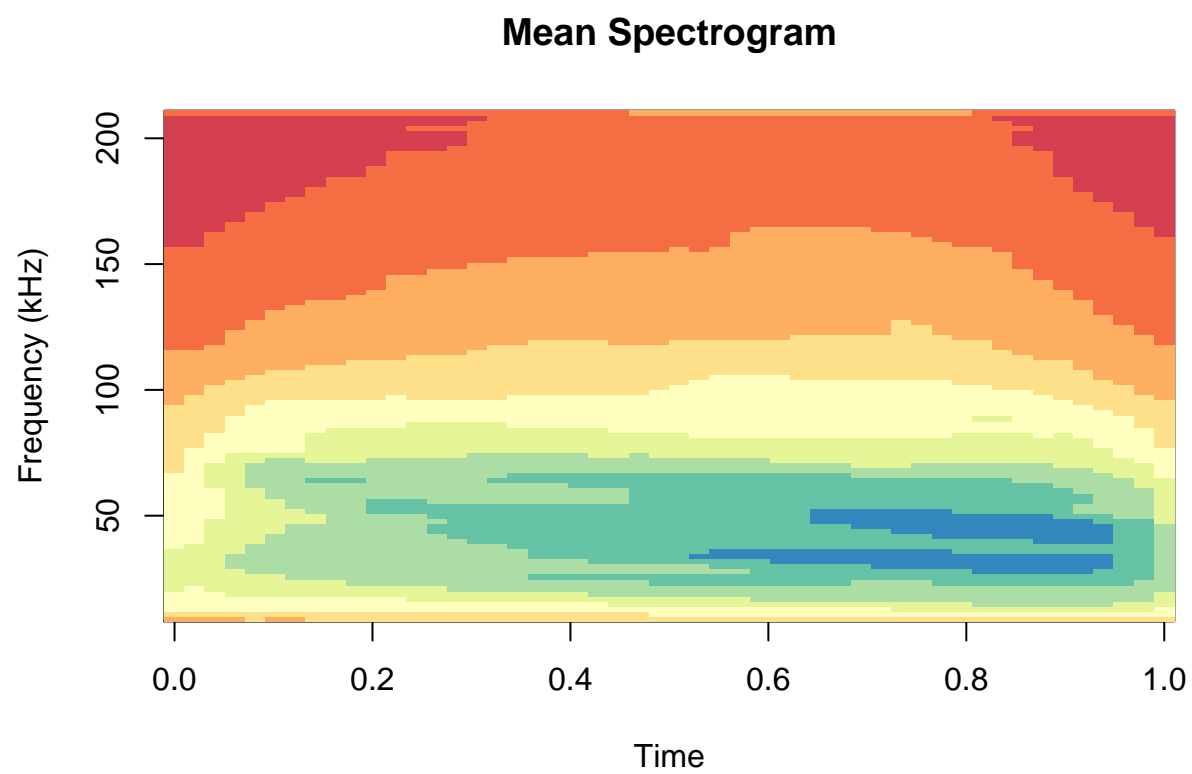


Figure 1: The mean spectrogram for the sample of echolocation calls

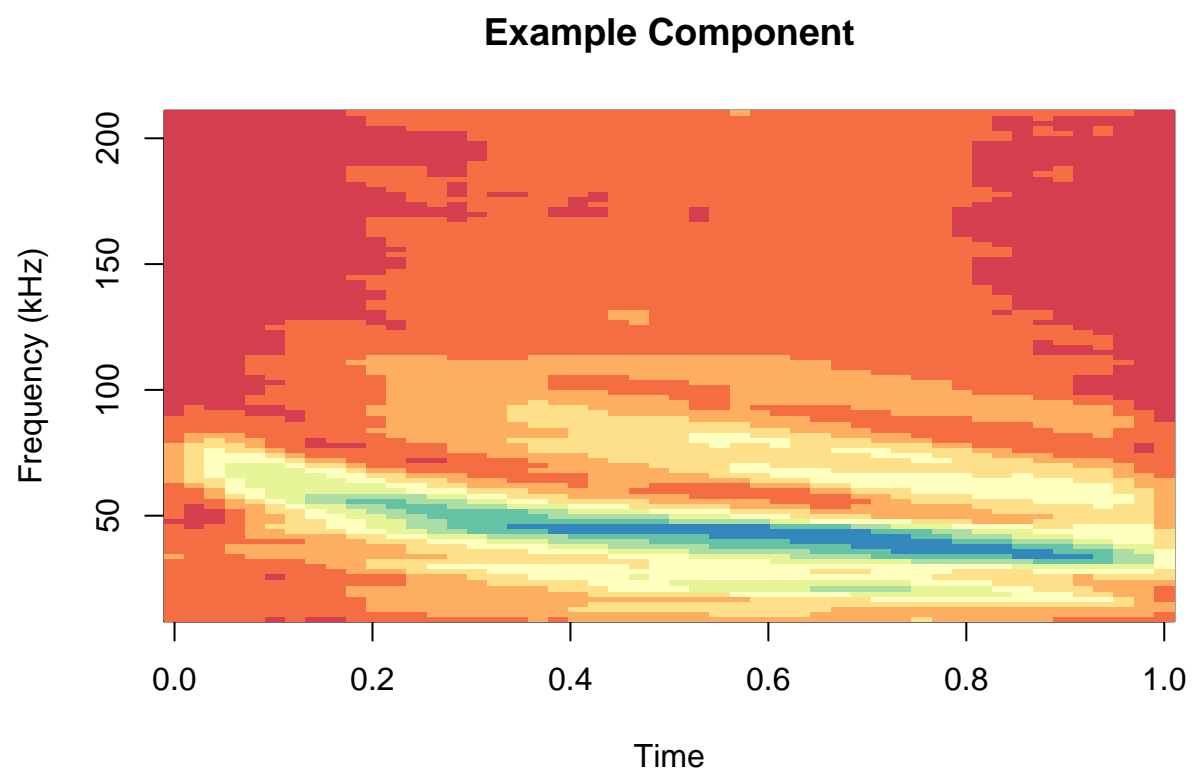


Figure 2: The first principal component of variation

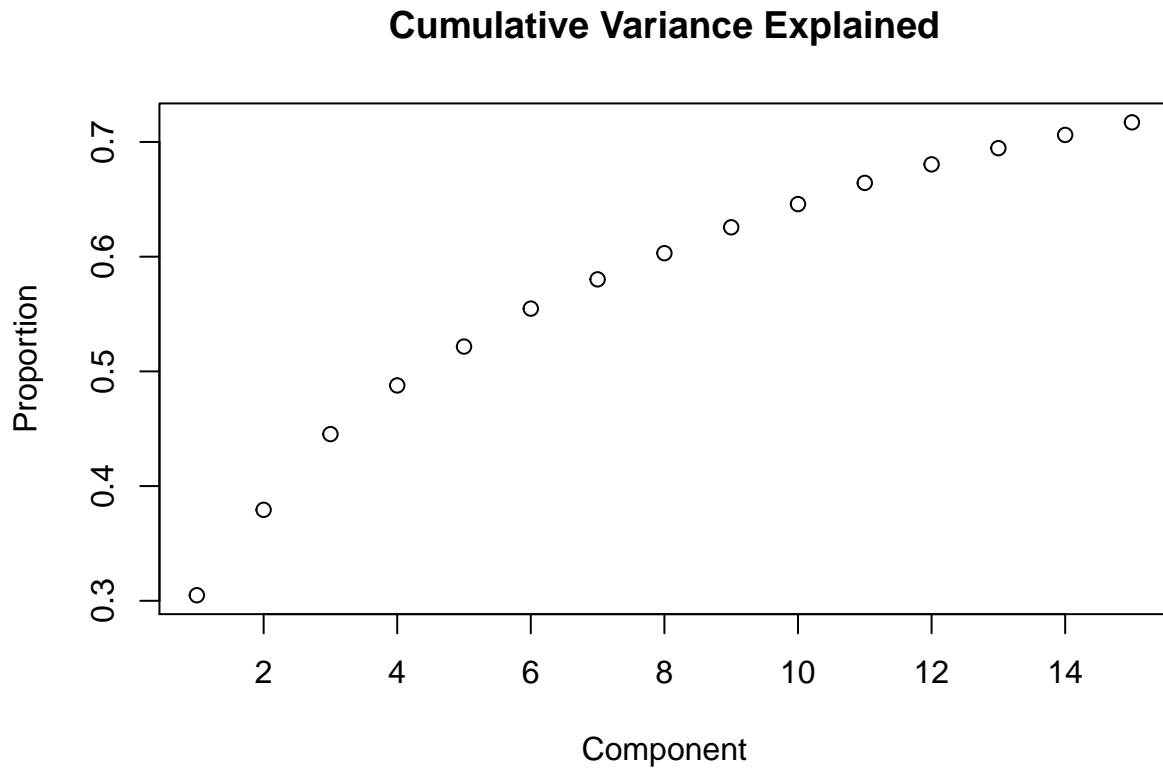


Figure 3: Plot demonstrating the proportion of variance explained by the first n components. The first 15 components explain approximately 71% of the variation. This is probably too many components to model and so perhaps only components which explain a proportion of the variance above a particular threshold.

Number of Components

Now that the mean spectrogram and principal components have been found, I can set about calculating component weights. But first the components to be considered must be identified.

```
variation %>%
  divide_by(variation %>% sum) %>%
  cumsum %>%
  extract(1:15) %>%
  plot(xlab = 'Component', ylab = 'Proportion',
       main = 'Cumulative Variance Explained')
```

```
variation %>%
  divide_by(variation %>% sum) %>%
  extract(1:15) %>%
  plot(xlab = 'Component', ylab = 'Proportion',
       main = 'Variance Explained')
abline(h = 0.025, col = 'red')
```

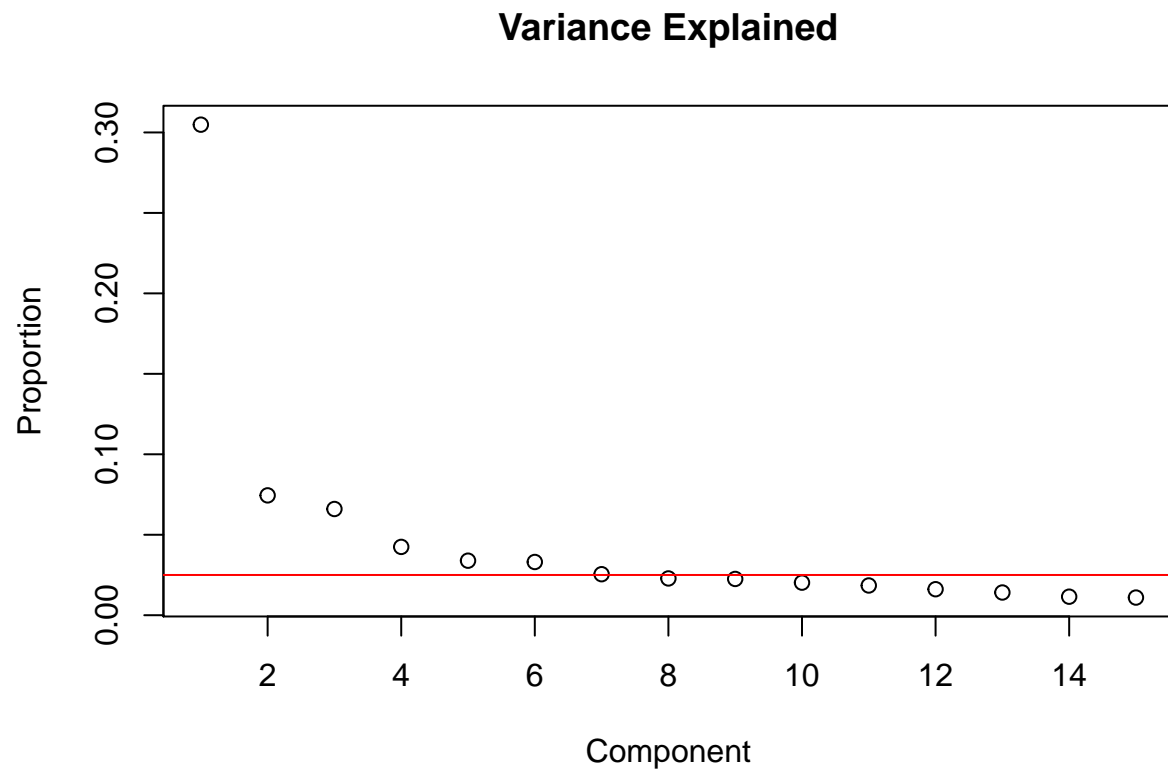


Figure 4: Plot demonstrating the proportion of variance explained by the n -th component. The red line represents a 2.5% threshold. 6 components lie above this threshold which corresponds to the number of components used in the study of spectral density curves.