Interpreting Canonical Correspondence Analysis in Ecology

Laxman Hegde

Department of Mathematics
Frostburg State University
Frostburg, Maryland 21532

lhegde2@frostburg.edu

April 27, 2012

Canonical Correspondence Analysis?

- Ter-Braak 1986: Read more details in this article. The Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis, **Ecology 67**.
- The Canonical Correspondence Analysis (CCA) is a multivariate method to elucidate the relationships between biological assemblages of species and their environment.
- The CCA is an extension of Correspondence Analysis (CA), also called reciprocal averaging.

Canonical Correspondence Analysis?

- Ter-Braak 1986: Read more details in this article. The Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis, **Ecology 67**.
- The Canonical Correspondence Analysis (CCA) is a multivariate method to elucidate the relationships between biological assemblages of species and their environment.
- The CCA is an extension of Correspondence Analysis (CA), also called reciprocal averaging.
- These methods are applications of Weighted Singular Value Decomposition (WSVD) which basically provides new coordinate bases (some rotations of standard basis) for the columns and rows of a matrix.
- Fundamentally speaking, these methods are based upon eigenvalue-eigenvector decomposition (spectral decomposition) methods of square symmetric matrices.

CCA: What is involved?

$\rightarrow$Data Step: **Y** species data matrix and **Z** environmental data matrix
$\rightarrow$**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
$\rightarrow$Weights for sites and species
$\rightarrow$Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
$\rightarrow$$\mathbf{F}_2$= relative frequencies for each site over all species

CCA: What is involved?

$\rightarrow$Data Step: **Y** species data matrix and **Z** environmental data matrix
$\rightarrow$**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
$\rightarrow$Weights for sites and species
$\rightarrow$Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
$\rightarrow$$\mathbf{F}_2$= relative frequencies for each site over all species
$\rightarrow$Standardize environmental variables and finding correlation matrix

CCA: What is involved?

→Data Step: **Y** species data matrix and **Z** environmental data matrix
→**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
→Weights for sites and species
→Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
→$\mathbf{F}_2$= relative frequencies for each site over all species
→Standardize environmental variables and finding correlation matrix
→**A** matrix = $\mathbf{F}_1\mathbf{Z}$ average environmental values for each species
→**W** matrix , weighted **A** matrix based on species weights and env correlations

CCA: What is involved?

$\rightarrow$Data Step: **Y** species data matrix and **Z** environmental data matrix
$\rightarrow$**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
$\rightarrow$Weights for sites and species
$\rightarrow$Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
$\rightarrow$$\mathbf{F}_2$= relative frequencies for each site over all species
$\rightarrow$Standardize environmental variables and finding correlation matrix
$\rightarrow$**A** matrix = $\mathbf{F}_1\mathbf{Z}$ average environmental values for each species
$\rightarrow$**W** matrix , weighted **A** matrix based on species weights and env correlations
$\rightarrow$Perform Singular Value Decomposition of **W** matrix
$\rightarrow$Interpret eigenvalues to see what percent of the total variation is explained by the first axes, 70 percent or higher is good.

CCA: What is involved?

$\rightarrow$Data Step: **Y** species data matrix and **Z** environmental data matrix
$\rightarrow$**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
$\rightarrow$Weights for sites and species
$\rightarrow$Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
$\rightarrow$$\mathbf{F}_2$= relative frequencies for each site over all species
$\rightarrow$Standardize environmental variables and finding correlation matrix
$\rightarrow$**A** matrix = $\mathbf{F}_1\mathbf{Z}$ average environmental values for each species
$\rightarrow$**W** matrix , weighted **A** matrix based on species weights and env correlations
$\rightarrow$Perform Singular Value Decomposition of **W** matrix
$\rightarrow$Interpret eigenvalues to see what percent of the total variation is explained by the first axes, 70 percent or higher is good.
$\rightarrow$Compute Canonical Coefficients or regression coefficients for environmental variables

CCA: What is involved?

→Data Step: **Y** species data matrix and **Z** environmental data matrix
→**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
→Weights for sites and species
→Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
→$\mathbf{F}_2$= relative frequencies for each site over all species
→Standardize environmental variables and finding correlation matrix
→**A** matrix = $\mathbf{F}_1\mathbf{Z}$ average environmental values for each species
→**W** matrix , weighted **A** matrix based on species weights and env correlations
→Perform Singular Value Decomposition of **W** matrix
→Interpret eigenvalues to see what percent of the total variation is explained by the first axes, 70 percent or higher is good.
→Compute Canonical Coefficients or regression coefficients for environmental variables
→Compute Species Scores and Sample Scores
→Compute Species-Environmental Correlations

CCA: What is involved?

→Data Step: **Y** species data matrix and **Z** environmental data matrix
→**F** matrix: Convert **Y** matrix to relative frequencies, dividing **Y** by sample size
→Weights for sites and species
→Preparing $\mathbf{F}_1$ and $\mathbf{F}_2$ matrices. $\mathbf{F}_1$= relative frequencies for each species over sites and
→$\mathbf{F}_2$= relative frequencies for each site over all species
→Standardize environmental variables and finding correlation matrix
→**A** matrix = $\mathbf{F}_1\mathbf{Z}$ average environmental values for each species
→**W** matrix , weighted **A** matrix based on species weights and env correlations
→Perform Singular Value Decomposition of **W** matrix
→Interpret eigenvalues to see what percent of the total variation is explained by the first axes, 70 percent or higher is good.
→Compute Canonical Coefficients or regression coefficients for environmental variables
→Compute Species Scores and Sample Scores
→Compute Species-Environmental Correlations
→Compute Biplot Coordinates for Species and Environmental axes
→ Generate Biplot picture and interpret results

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.
- Note again that it is possible to add more analysis to CCA other than the basic analysis we did in our program.

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.
- Note again that it is possible to add more analysis to CCA other than the basic analysis we did in our program.
- The users of this program should be familiar with basics of SAS and be able to make necessary changes to meet the individual research interests such as reading data sets, printing and managing output delivery system (ODS).

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.
- Note again that it is possible to add more analysis to CCA other than the basic analysis we did in our program.
- The users of this program should be familiar with basics of SAS and be able to make necessary changes to meet the individual research interests such as reading data sets, printing and managing output delivery system (ODS).
- In this presentation, we try to provide some basic interpretations to some important results of CCA.

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.
- Note again that it is possible to add more analysis to CCA other than the basic analysis we did in our program.
- The users of this program should be familiar with basics of SAS and be able to make necessary changes to meet the individual research interests such as reading data sets, printing and managing output delivery system (ODS).
- In this presentation, we try to provide some basic interpretations to some important results of CCA.
- Although it is not the only goal of our program, the biplot is an important output. In individual cases, the users may have to adjust scaling factors in order to receive a more visible picture.

SAS PROGRAM

- Based upon Ter-Braak (1986), we published our paper SUGI 24(1999) and a SAS program.
- The SAS program analyzes Hunting birds data as an example to demonstrate some important results of CCA.
- Note again that it is possible to add more analysis to CCA other than the basic analysis we did in our program.
- The users of this program should be familiar with basics of SAS and be able to make necessary changes to meet the individual research interests such as reading data sets, printing and managing output delivery system (ODS).
- In this presentation, we try to provide some basic interpretations to some important results of CCA.
- Although it is not the only goal of our program, the biplot is an important output. In individual cases, the users may have to adjust scaling factors in order to receive a more visible picture.
- Interpreting a biplot requires special skills. If you are willing to share your data set, we will try to provide some assistance in this regard.

Some basic matrices in our SAS Program

Basic Data:

- **Y** : Species data, n sites by m species matrix.
- **Z** : Environmental data, n sites by q variables matrix. This is generally standardized to have mean zero and variance 1 to remove the effects of varying scales of measurements.
- **F** : We analyze this matrix of relative frequencies instead of original counts **Y**.

Some basic matrices in our SAS Program

Basic Data:

- **Y** : Species data, n sites by m species matrix.
- **Z** : Environmental data, n sites by q variables matrix. This is generally standardized to have mean zero and variance 1 to remove the effects of varying scales of measurements.
- **F** : We analyze this matrix of relative frequencies instead of original counts **Y**.

Row and Column Profiles for Species:

- $\mathbf{F}_1$: Relative frequency distribution of species over sites, m by n matrix.
- $\mathbf{F}_2$: Relative frequency distribution of sites over species, n by m matrix.
- These matrices are related to each other through species and site weights.Hence, some use the word correspondence because of the correspondence between these two profiles.

Some basic matrices in our SAS Program

Basic Data:

- **Y** : Species data, n sites by m species matrix.
- **Z** : Environmental data, n sites by q variables matrix. This is generally standardized to have mean zero and variance 1 to remove the effects of varying scales of measurements.
- **F** : We analyze this matrix of relative frequencies instead of original counts **Y**.

Row and Column Profiles for Species:

- $\mathbf{F}_1$: Relative frequency distribution of species over sites, m by n matrix.
- $\mathbf{F}_2$: Relative frequency distribution of sites over species, n by m matrix.
- These matrices are related to each other through species and site weights.Hence, some use the word correspondence because of the correspondence between these two profiles.

Weight Matrices

- Species Weights, m by m diagonal matrix.
- Site Weights, n by n diagonal matrix.
- Observed correlation matrix of environmental variables, q by q matrix.

Some basic matrices in our SAS Program

Basic Data:

- **Y** : Species data, n sites by m species matrix.
- **Z** : Environmental data, n sites by q variables matrix. This is generally standardized to have mean zero and variance 1 to remove the effects of varying scales of measurements.
- **F** : We analyze this matrix of relative frequencies instead of original counts **Y**.

Row and Column Profiles for Species:

- $\mathbf{F}_1$: Relative frequency distribution of species over sites, m by n matrix.
- $\mathbf{F}_2$: Relative frequency distribution of sites over species, n by m matrix.
- These matrices are related to each other through species and site weights.Hence, some use the word correspondence because of the correspondence between these two profiles.

Weight Matrices

- Species Weights, m by m diagonal matrix.
- Site Weights, n by n diagonal matrix.
- Observed correlation matrix of environmental variables, q by q matrix.

It may be of interest to some researchers to do some descriptive statistics/exploratory analysis using SAS procedures such as PROC UNIVARIATE, PROC MEANS, PROC CORR, and some other graphs using PROC SGPLOT.

Data Matrices: Y Species Data, Z Env Data

$$
\mathbf{Y} = \begin{pmatrix}
0 & 2 & 1 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 1 & 1 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 \\
0 & 3 & 1 & 0 & 0 & 0 & 4 & 1 & 0 & 0 & 0 & 0 \\
0 & 2 & 2 & 1 & 0 & 0 & 5 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 5 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 3 & 3 & 6 & 5 & 8 & 1 & 1 & 0 & 0 & 0 \\
0 & 7 & 1 & 1 & 1 & 2 & 5 & 3 & 1 & 0 & 0 & 0 \\
0 & 4 & 1 & 0 & 1 & 0 & 4 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 4 & 9 & 8 & 3 & 9 & 4 & 1 & 1 & 0 & 0 \\
2 & 0 & 5 & 5 & 4 & 2 & 7 & 2 & 3 & 0 & 0 & 0 \\
1 & 1 & 5 & 3 & 8 & 2 & 9 & 1 & 3 & 0 & 0 & 0 \\
1 & 1 & 5 & 5 & 9 & 4 & 9 & 2 & 2 & 1 & 0 & 0 \\
3 & 1 & 4 & 9 & 9 & 4 & 9 & 2 & 5 & 1 & 0 & 0 \\
1 & 1 & 4 & 7 & 8 & 4 & 9 & 6 & 4 & 1 & 1 & 0 \\
1 & 1 & 1 & 4 & 6 & 3 & 8 & 4 & 5 & 3 & 1 & 0 \\
0 & 0 & 2 & 3 & 6 & 2 & 7 & 3 & 7 & 5 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 5 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 2 & 0 & 3 & 3 & 9 & 4 & 0 & 0 \\
0 & 1 & 2 & 2 & 0 & 1 & 4 & 1 & 3 & 3 & 3 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 2 & 1 & 9 & 3 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 4 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 3 & 3 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 4 & 3 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 5 & 3 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 4 & 2 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 2 & 4
\end{pmatrix}
$$

$$
\mathbf{Z} = \begin{pmatrix}
9 & 0 & 1 & 1 & 9 & 5 \\
7 & 0 & 3 & 0 & 9 & 2 \\
8 & 0 & 1 & 0 & 9 & 0 \\
8 & 0 & 1 & 0 & 9 & 0 \\
9 & 0 & 1 & 2 & 9 & 5 \\
8 & 0 & 0 & 2 & 9 & 5 \\
8 & 0 & 2 & 3 & 3 & 9 \\
6 & 0 & 2 & 1 & 9 & 6 \\
7 & 0 & 1 & 0 & 9 & 2 \\
8 & 0 & 0 & 5 & 0 & 9 \\
9 & 5 & 5 & 1 & 7 & 6 \\
8 & 0 & 4 & 2 & 0 & 9 \\
6 & 0 & 5 & 6 & 0 & 9 \\
8 & 0 & 1 & 5 & 0 & 9 \\
9 & 3 & 1 & 7 & 3 & 9 \\
6 & 0 & 5 & 8 & 0 & 9 \\
5 & 0 & 7 & 8 & 0 & 9 \\
5 & 0 & 9 & 7 & 0 & 6 \\
6 & 0 & 8 & 8 & 0 & 8 \\
3 & 7 & 2 & 5 & 0 & 8 \\
4 & 0 & 9 & 8 & 0 & 7 \\
4 & 8 & 7 & 8 & 0 & 5 \\
0 & 7 & 8 & 8 & 0 & 6 \\
0 & 6 & 9 & 9 & 0 & 6 \\
1 & 7 & 9 & 8 & 0 & 0 \\
0 & 5 & 8 & 8 & 0 & 6 \\
2 & 7 & 9 & 9 & 0 & 5 \\
0 & 9 & 4 & 9 & 0 & 2
\end{pmatrix}
$$

Note: **Y** is 28 by 12 matrix and **Z** is 28 by 6 matrix. In CA, reciprocal averaging, we analyze only **Y** matrix assuming there is only one not measured environmental variable, say altitude of the sites.

Relative Frequencies of Species Matrix Y

All numbers add to 1

$$
\mathbf{F} = \begin{pmatrix}
0. & 0.004 & 0.002 & 0. & 0. & 0. & 0.009 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.005 & 0.002 & 0.002 & 0. & 0. & 0.007 & 0.002 & 0. & 0. & 0. & 0. \\
0. & 0.005 & 0.002 & 0. & 0. & 0. & 0.007 & 0.002 & 0. & 0. & 0. & 0. \\
0. & 0.004 & 0.004 & 0.002 & 0. & 0. & 0.009 & 0.002 & 0. & 0. & 0. & 0. \\
0. & 0.002 & 0.002 & 0. & 0. & 0. & 0.007 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.004 & 0. & 0. & 0. & 0. & 0.009 & 0.002 & 0. & 0. & 0. & 0. \\
0. & 0.002 & 0.005 & 0.005 & 0.011 & 0.009 & 0.014 & 0.002 & 0.002 & 0. & 0. & 0. \\
0. & 0.013 & 0.002 & 0.002 & 0.002 & 0.004 & 0.009 & 0.005 & 0.002 & 0. & 0. & 0. \\
0. & 0.007 & 0.002 & 0. & 0.002 & 0. & 0.007 & 0.002 & 0.002 & 0. & 0. & 0. \\
0.002 & 0.002 & 0.007 & 0.016 & 0.014 & 0.005 & 0.016 & 0.007 & 0.002 & 0.002 & 0. & 0. \\
0.004 & 0. & 0.009 & 0.009 & 0.007 & 0.004 & 0.013 & 0.004 & 0.005 & 0. & 0. & 0. \\
0.002 & 0.002 & 0.009 & 0.005 & 0.014 & 0.004 & 0.016 & 0.002 & 0.005 & 0. & 0. & 0. \\
0.002 & 0.002 & 0.009 & 0.009 & 0.016 & 0.007 & 0.016 & 0.004 & 0.004 & 0.002 & 0. & 0. \\
0.005 & 0.002 & 0.007 & 0.016 & 0.016 & 0.007 & 0.016 & 0.004 & 0.009 & 0.002 & 0. & 0. \\
0.002 & 0.002 & 0.007 & 0.013 & 0.014 & 0.007 & 0.016 & 0.011 & 0.007 & 0.002 & 0.002 & 0. \\
0.002 & 0.002 & 0.002 & 0.007 & 0.011 & 0.005 & 0.014 & 0.007 & 0.009 & 0.005 & 0.002 & 0. \\
0. & 0. & 0.004 & 0.005 & 0.011 & 0.004 & 0.013 & 0.005 & 0.013 & 0.009 & 0. & 0. \\
0. & 0. & 0. & 0.002 & 0.002 & 0. & 0.002 & 0.002 & 0.009 & 0.002 & 0. & 0. \\
0. & 0. & 0. & 0.002 & 0.004 & 0. & 0.005 & 0.005 & 0.016 & 0.007 & 0. & 0. \\
0. & 0.002 & 0.004 & 0.004 & 0. & 0.002 & 0.007 & 0.002 & 0.005 & 0.005 & 0.005 & 0. \\
0. & 0. & 0. & 0. & 0.002 & 0.002 & 0.004 & 0.002 & 0.016 & 0.005 & 0.002 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0. & 0.007 & 0.002 & 0.002 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0. & 0.004 & 0.005 & 0.005 & 0.002 \\
0. & 0.002 & 0. & 0. & 0. & 0. & 0.002 & 0. & 0.004 & 0.007 & 0.005 & 0.004 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0. & 0.002 & 0.004 & 0.007 & 0.002 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0.009 & 0.005 & 0.004 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0.005 & 0.007 & 0.004 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.002 & 0. & 0. & 0.002 & 0.004 & 0.007
\end{pmatrix}
$$

Note: In CA, we analyze this matrix and find site scores and species scores.

Relative Frequency Distribution of Species over Sites

Each Column adds to 1

$$
\mathbf{F}_1' = \begin{pmatrix}
0. & 0.061 & 0.023 & 0. & 0. & 0. & 0.04 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.091 & 0.023 & 0.018 & 0. & 0. & 0.032 & 0.026 & 0. & 0. & 0. & 0. \\
0. & 0.091 & 0.023 & 0. & 0. & 0. & 0.032 & 0.026 & 0. & 0. & 0. & 0. \\
0. & 0.061 & 0.047 & 0.018 & 0. & 0. & 0.04 & 0.026 & 0. & 0. & 0. & 0. \\
0. & 0.03 & 0.023 & 0. & 0. & 0. & 0.032 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.061 & 0. & 0. & 0. & 0. & 0.04 & 0.026 & 0. & 0. & 0. & 0. \\
0. & 0.03 & 0.07 & 0.055 & 0.086 & 0.152 & 0.063 & 0.026 & 0.014 & 0. & 0. & 0. \\
0. & 0.212 & 0.023 & 0.018 & 0.014 & 0.061 & 0.04 & 0.077 & 0.014 & 0. & 0. & 0. \\
0. & 0.121 & 0.023 & 0. & 0.014 & 0. & 0.032 & 0.026 & 0.014 & 0. & 0. & 0. \\
0.1 & 0.03 & 0.093 & 0.164 & 0.114 & 0.091 & 0.071 & 0.103 & 0.014 & 0.024 & 0. & 0. \\
0.2 & 0. & 0.116 & 0.091 & 0.057 & 0.061 & 0.056 & 0.051 & 0.043 & 0. & 0. & 0. \\
0.1 & 0.03 & 0.116 & 0.055 & 0.114 & 0.061 & 0.071 & 0.026 & 0.043 & 0. & 0. & 0. \\
0.1 & 0.03 & 0.116 & 0.091 & 0.129 & 0.121 & 0.071 & 0.051 & 0.029 & 0.024 & 0. & 0. \\
0.3 & 0.03 & 0.093 & 0.164 & 0.129 & 0.121 & 0.071 & 0.051 & 0.071 & 0.024 & 0. & 0. \\
0.1 & 0.03 & 0.093 & 0.127 & 0.114 & 0.121 & 0.071 & 0.154 & 0.057 & 0.024 & 0.038 & 0. \\
0.1 & 0.03 & 0.023 & 0.073 & 0.086 & 0.091 & 0.063 & 0.103 & 0.071 & 0.071 & 0.038 & 0. \\
0. & 0. & 0.047 & 0.055 & 0.086 & 0.061 & 0.056 & 0.077 & 0.1 & 0.119 & 0. & 0. \\
0. & 0. & 0. & 0.018 & 0.014 & 0. & 0.008 & 0.026 & 0.071 & 0.024 & 0. & 0. \\
0. & 0. & 0. & 0.018 & 0.029 & 0. & 0.024 & 0.077 & 0.129 & 0.095 & 0. & 0. \\
0. & 0.03 & 0.047 & 0.036 & 0. & 0.03 & 0.032 & 0.026 & 0.043 & 0.071 & 0.115 & 0. \\
0. & 0. & 0. & 0. & 0.014 & 0.03 & 0.016 & 0.026 & 0.129 & 0.071 & 0.038 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.008 & 0. & 0.057 & 0.024 & 0.038 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.008 & 0. & 0.029 & 0.071 & 0.115 & 0.083 \\
0. & 0.03 & 0. & 0. & 0. & 0. & 0.008 & 0. & 0.029 & 0.095 & 0.115 & 0.167 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.008 & 0. & 0.014 & 0.048 & 0.154 & 0.083 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.014 & 0.119 & 0.115 & 0.167 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.014 & 0.071 & 0.154 & 0.167 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.008 & 0. & 0. & 0.024 & 0.077 & 0.333
\end{pmatrix}
$$

Relative Frequency Distribution of Species in a site

Each Row adds to 1

$$
\mathbf{F}_2 = \begin{pmatrix}
0. & 0.25 & 0.125 & 0. & 0. & 0. & 0.625 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.3 & 0.1 & 0.1 & 0. & 0. & 0.4 & 0.1 & 0. & 0. & 0. & 0. \\
0. & 0.333 & 0.111 & 0. & 0. & 0. & 0.444 & 0.111 & 0. & 0. & 0. & 0. \\
0. & 0.182 & 0.182 & 0.091 & 0. & 0. & 0.455 & 0.091 & 0. & 0. & 0. & 0. \\
0. & 0.167 & 0.167 & 0. & 0. & 0. & 0.667 & 0. & 0. & 0. & 0. & 0. \\
0. & 0.25 & 0. & 0. & 0. & 0. & 0.625 & 0.125 & 0. & 0. & 0. & 0. \\
0. & 0.036 & 0.107 & 0.107 & 0.214 & 0.179 & 0.286 & 0.036 & 0.036 & 0. & 0. & 0. \\
0. & 0.333 & 0.048 & 0.048 & 0.048 & 0.095 & 0.238 & 0.143 & 0.048 & 0. & 0. & 0. \\
0. & 0.333 & 0.083 & 0. & 0.083 & 0. & 0.333 & 0.083 & 0.083 & 0. & 0. & 0. \\
0.024 & 0.024 & 0.098 & 0.22 & 0.195 & 0.073 & 0.22 & 0.098 & 0.024 & 0.024 & 0. & 0. \\
0.067 & 0. & 0.167 & 0.167 & 0.133 & 0.067 & 0.233 & 0.067 & 0.1 & 0. & 0. & 0. \\
0.03 & 0.03 & 0.152 & 0.091 & 0.242 & 0.061 & 0.273 & 0.03 & 0.091 & 0. & 0. & 0. \\
0.026 & 0.026 & 0.128 & 0.128 & 0.231 & 0.103 & 0.231 & 0.051 & 0.051 & 0.026 & 0. & 0. \\
0.064 & 0.021 & 0.085 & 0.191 & 0.191 & 0.085 & 0.191 & 0.043 & 0.106 & 0.021 & 0. & 0. \\
0.022 & 0.022 & 0.087 & 0.152 & 0.174 & 0.087 & 0.196 & 0.13 & 0.087 & 0.022 & 0.022 & 0. \\
0.027 & 0.027 & 0.027 & 0.108 & 0.162 & 0.081 & 0.216 & 0.108 & 0.135 & 0.081 & 0.027 & 0. \\
0. & 0. & 0.057 & 0.086 & 0.171 & 0.057 & 0.2 & 0.086 & 0.2 & 0.143 & 0. & 0. \\
0. & 0. & 0. & 0.1 & 0.1 & 0. & 0.1 & 0.1 & 0.5 & 0.1 & 0. & 0. \\
0. & 0. & 0. & 0.045 & 0.091 & 0. & 0.136 & 0.136 & 0.409 & 0.182 & 0. & 0. \\
0. & 0.05 & 0.1 & 0.1 & 0. & 0.05 & 0.2 & 0.05 & 0.15 & 0.15 & 0.15 & 0. \\
0. & 0. & 0. & 0. & 0.056 & 0.056 & 0.111 & 0.056 & 0.5 & 0.167 & 0.056 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.143 & 0. & 0.571 & 0.143 & 0.143 & 0. \\
0. & 0. & 0. & 0. & 0. & 0. & 0.1 & 0. & 0.2 & 0.3 & 0.3 & 0.1 \\
0. & 0.077 & 0. & 0. & 0. & 0. & 0.077 & 0. & 0.154 & 0.308 & 0.231 & 0.154 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.111 & 0. & 0.111 & 0.222 & 0.444 & 0.111 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.091 & 0.455 & 0.273 & 0.182 \\
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0.1 & 0.3 & 0.4 & 0.2 \\
0. & 0. & 0. & 0. & 0. & 0. & 0.125 & 0. & 0. & 0.125 & 0.25 & 0.5
\end{pmatrix}
$$

Weight Matrices

$$SpeciesWeights = \begin{pmatrix} 0.018 \\ 0.059 \\ 0.077 \\ 0.098 \\ 0.125 \\ 0.059 \\ 0.225 \\ 0.07 \\ 0.125 \\ 0.075 \\ 0.047 \\ 0.021 \end{pmatrix}$$

Note: We put this info in a diagonal matrix (12 by 12) and call it $\mathbf{D}_c$.

$$SiteWeights = \begin{pmatrix} 0.014 \\ 0.018 \\ 0.016 \\ 0.02 \\ 0.011 \\ 0.014 \\ 0.05 \\ 0.038 \\ 0.021 \\ 0.073 \\ 0.054 \\ 0.059 \\ 0.07 \\ 0.084 \\ 0.082 \\ 0.066 \\ 0.063 \\ 0.018 \\ 0.039 \\ 0.036 \\ 0.032 \\ 0.013 \\ 0.018 \\ 0.023 \\ 0.016 \\ 0.02 \\ 0.018 \\ 0.014 \end{pmatrix}$$

This info is used in computing the means, variances, and the correlations of environmental variables.

Standardized Environmental Variables, mean =0 and standard deviation =1

$$
\mathbf{z}_{standard} = \begin{pmatrix}
1.0562 & -0.597975 & -0.970245 & -1.44087 & 2.01791 & -0.871263 \\
0.280868 & -0.597975 & -0.300656 & -1.7888 & 2.01791 & -2.0439 \\
0.668536 & -0.597975 & -0.970245 & -1.7888 & 2.01791 & -2.82566 \\
0.668536 & -0.597975 & -0.970245 & -1.7888 & 2.01791 & -2.82566 \\
1.0562 & -0.597975 & -0.970245 & -1.09295 & 2.01791 & -0.871263 \\
0.668536 & -0.597975 & -1.30504 & -1.09295 & 2.01791 & -0.871263 \\
0.668536 & -0.597975 & -0.63545 & -0.745022 & 0.252634 & 0.692256 \\
-0.106799 & -0.597975 & -0.63545 & -1.44087 & 2.01791 & -0.480383 \\
0.280868 & -0.597975 & -0.970245 & -1.7888 & 2.01791 & -2.0439 \\
0.668536 & -0.597975 & -1.30504 & -0.0491702 & -0.630006 & 0.692256 \\
1.0562 & 1.27572 & 0.368933 & -1.44087 & 1.42949 & -0.480383 \\
0.668536 & -0.597975 & 0.0341382 & -1.09295 & -0.630006 & 0.692256 \\
-0.106799 & -0.597975 & 0.368933 & 0.298756 & -0.630006 & 0.692256 \\
0.668536 & -0.597975 & -0.970245 & -0.0491702 & -0.630006 & 0.692256 \\
1.0562 & 0.526245 & -0.970245 & 0.646681 & 0.252634 & 0.692256 \\
-0.106799 & -0.597975 & 0.368933 & 0.994607 & -0.630006 & 0.692256 \\
-0.494467 & -0.597975 & 1.03852 & 0.994607 & -0.630006 & 0.692256 \\
-0.494467 & -0.597975 & 1.70811 & 0.646681 & -0.630006 & -0.480383 \\
-0.106799 & -0.597975 & 1.37332 & 0.994607 & -0.630006 & 0.301376 \\
-1.2698 & 2.0252 & -0.63545 & -0.0491702 & -0.630006 & 0.301376 \\
-0.882134 & -0.597975 & 1.70811 & 0.994607 & -0.630006 & -0.0895037 \\
-0.882134 & 2.39994 & 1.03852 & 0.994607 & -0.630006 & -0.871263 \\
-2.4328 & 2.0252 & 1.37332 & 0.994607 & -0.630006 & -0.480383 \\
-2.4328 & 1.65046 & 1.70811 & 1.34253 & -0.630006 & -0.480383 \\
-2.04514 & 2.0252 & 1.70811 & 0.994607 & -0.630006 & -2.82566 \\
-2.4328 & 1.27572 & 1.37332 & 0.994607 & -0.630006 & -0.480383 \\
-1.65747 & 2.0252 & 1.70811 & 1.34253 & -0.630006 & -0.871263 \\
-2.4328 & 2.77468 & 0.0341382 & 1.34253 & -0.630006 & -2.0439
\end{pmatrix}
$$

Correlation Matrix of Observed Env Variables

$$
\mathit{Corrmatrix} = \begin{pmatrix}
1. & -0.56 & -0.688 & -0.573 & 0.401 & 0.266 \\
-0.56 & 1. & 0.286 & 0.27 & -0.088 & -0.312 \\
-0.688 & 0.286 & 1. & 0.558 & -0.396 & -0.099 \\
-0.573 & 0.27 & 0.558 & 1. & -0.772 & 0.348 \\
0.401 & -0.088 & -0.396 & -0.772 & 1. & -0.592 \\
0.266 & -0.312 & -0.099 & 0.348 & -0.592 & 1.
\end{pmatrix}
$$

$$
\mathit{EnvWeightMatrix} = \begin{pmatrix}
3.761 & 0.917 & 1.194 & 0.802 & -1.315 & -1.654 \\
0.917 & 1.581 & 0.309 & -0.31 & -0.178 & 0.282 \\
1.194 & 0.309 & 2.127 & -0.642 & 0.033 & 0.232 \\
0.802 & -0.31 & -0.642 & 3.405 & 1.698 & -0.553 \\
-1.315 & -0.178 & 0.033 & 1.698 & 4.097 & 2.132 \\
-1.654 & 0.282 & 0.232 & -0.553 & 2.132 & 3.006
\end{pmatrix}
$$

Average Environmental Values for Species of CCA

A= F_1Z

$$\mathbf{A} = \begin{pmatrix} 0.629769 & -0.110813 & -0.367615 & -0.223133 & -0.129843 & 0.457728 \\ 0.269121 & -0.416283 & -0.655741 & -1.10349 & 1.34925 & -0.954177 \\ 0.47921 & -0.153516 & -0.386301 & -0.494192 & 0.245792 & 0.0650301 \\ 0.435935 & -0.189168 & -0.379789 & -0.0744739 & -0.137867 & 0.407979 \\ 0.39163 & -0.362424 & -0.181087 & 0.00550386 & -0.260138 & 0.53032 \\ 0.351353 & -0.268658 & -0.290511 & -0.0702566 & -0.103988 & 0.514583 \\ 0.299329 & -0.220261 & -0.297999 & -0.372244 & 0.360045 & -0.142241 \\ 0.280868 & -0.26167 & -0.180473 & -0.040249 & 0.177195 & 0.0909022 \\ -0.350476 & 0.0712035 & 0.689378 & 0.517452 & -0.403042 & 0.0891841 \\ -1.1775 & 0.704692 & 0.982722 & 0.903484 & -0.608991 & -0.173264 \\ -1.77675 & 1.70812 & 1.0514 & 0.981226 & -0.596058 & -0.841195 \\ -2.27128 & 2.08766 & 1.06642 & 1.22656 & -0.630006 & -1.26214 \end{pmatrix}$$

- Each row corresponds to a species and each column corresponds to a standardized environmental variable.
- Row Cloud: 12 species, each with 6 coordinates and Column Cloud: 6 variables,each with 12 coordinates.

Average Environmental Values for Species of CCA

$A = F_1 Z$

$$\mathbf{A} = \begin{pmatrix} 0.629769 & -0.110813 & -0.367615 & -0.223133 & -0.129843 & 0.457728 \\ 0.269121 & -0.416283 & -0.655741 & -1.10349 & 1.34925 & -0.954177 \\ 0.47921 & -0.153516 & -0.386301 & -0.494192 & 0.245792 & 0.0650301 \\ 0.435935 & -0.189168 & -0.379789 & -0.0744739 & -0.137867 & 0.407979 \\ 0.39163 & -0.362424 & -0.181087 & 0.00550386 & -0.260138 & 0.53032 \\ 0.351353 & -0.268658 & -0.290511 & -0.0702566 & -0.103988 & 0.514583 \\ 0.299329 & -0.220261 & -0.297999 & -0.372244 & 0.360045 & -0.142241 \\ 0.280868 & -0.26167 & -0.180473 & -0.040249 & 0.177195 & 0.0909022 \\ -0.350476 & 0.0712035 & 0.689378 & 0.517452 & -0.403042 & 0.0891841 \\ -1.1775 & 0.704692 & 0.982722 & 0.903484 & -0.608991 & -0.173264 \\ -1.77675 & 1.70812 & 1.0514 & 0.981226 & -0.596058 & -0.841195 \\ -2.27128 & 2.08766 & 1.06642 & 1.22656 & -0.630006 & -1.26214 \end{pmatrix}$$

- Each row corresponds to a species and each column corresponds to a standardized environmental variable.
- Row Cloud: 12 species, each with 6 coordinates and Column Cloud: 6 variables,each with 12 coordinates.
- Each entry in this matrix is an average standardized environmental value for a given species.
- Note the signs. For example, in the first column, the last four values are negative indicating those species prefer dryness if we interpret the first column as water.

Average Environmental Values for Species of CCA

A= $\mathbf{F}_1\mathbf{Z}$

$$
\mathbf{A} = \begin{pmatrix}
0.629769 & -0.110813 & -0.367615 & -0.223133 & -0.129843 & 0.457728 \\
0.269121 & -0.416283 & -0.655741 & -1.10349 & 1.34925 & -0.954177 \\
0.47921 & -0.153516 & -0.386301 & -0.494192 & 0.245792 & 0.0650301 \\
0.435935 & -0.189168 & -0.379789 & -0.0744739 & -0.137867 & 0.407979 \\
0.39163 & -0.362424 & -0.181087 & 0.00550386 & -0.260138 & 0.53032 \\
0.351353 & -0.268658 & -0.290511 & -0.0702566 & -0.103988 & 0.514583 \\
0.299329 & -0.220261 & -0.297999 & -0.372244 & 0.360045 & -0.142241 \\
0.280868 & -0.26167 & -0.180473 & -0.040249 & 0.177195 & 0.0909022 \\
-0.350476 & 0.0712035 & 0.689378 & 0.517452 & -0.403042 & 0.0891841 \\
-1.1775 & 0.704692 & 0.982722 & 0.903484 & -0.608991 & -0.173264 \\
-1.77675 & 1.70812 & 1.0514 & 0.981226 & -0.596058 & -0.841195 \\
-2.27128 & 2.08766 & 1.06642 & 1.22656 & -0.630006 & -1.26214
\end{pmatrix}
$$

- Each row corresponds to a species and each column corresponds to a standardized environmental variable.
- Row Cloud: 12 species, each with 6 coordinates and Column Cloud: 6 variables,each with 12 coordinates.
- Each entry in this matrix is an average standardized environmental value for a given species.
- Note the signs. For example, in the first column, the last four values are negative indicating those species prefer dryness if we interpret the first column as water.
- CCA in ecology analyzes this matrix given proper weights for species (rows) and proper weights for environmental variables (columns). Generally, species weights are frequencies or relative frequencies of their occurrences and env weights are in the inverse correlation matrix of **Z**.

The Weighted **A** matrix, called **W** in our program

$\mathbf{W} = SpeWt^{1/2}\mathbf{A}Corr^{-1/2}$: A mathematical requirement in Weighted Analysis

$$\mathbf{W} = \begin{pmatrix} 0.0905928 & 0.0221424 & -0.0188277 & -0.0339005 & -0.042086 & 0.043711 \\ -0.0790803 & -0.120468 & -0.122306 & -0.144538 & 0.262412 & -0.147858 \\ 0.0985968 & 0.00620758 & -0.048164 & -0.124552 & 0.00421431 & 0.02378 \\ 0.118231 & -0.00622075 & -0.0904924 & -0.0130824 & -0.0654825 & 0.0897482 \\ 0.11948 & -0.0775277 & -0.0222739 & -0.0163238 & -0.0955793 & 0.138474 \\ 0.0441676 & -0.0301789 & -0.0471713 & -0.0191927 & -0.00946398 & 0.122513 \\ 0.0698714 & -0.0782216 & -0.0843257 & -0.095609 & 0.109338 & -0.0502717 \\ 0.05573 & -0.0562631 & -0.0272998 & 0.0460913 & 0.0662339 & 0.0193313 \\ -0.0126899 & -0.0128693 & 0.224758 & 0.119222 & -0.072703 & 0.00209805 \\ -0.229019 & 0.10246 & 0.155209 & 0.145066 & -0.0648578 & -0.0300013 \\ -0.257685 & 0.27168 & 0.0779809 & 0.107134 & -0.0885281 & -0.151806 \\ -0.241603 & 0.212812 & 0.0113994 & 0.115348 & -0.0644476 & -0.166911 \end{pmatrix}$$

- This is weighted **A** matrix. Interpreting this matrix mat not be simple unlike **A** matrix.
- For example, the first element 0.0905928 in **W** is a result of the inner product of the first row of **A** and the first column of inverse of square root of the correlation matrix. Then this inner product gets scaled by the square root of the species 1 weight.
- The users who do not have matrix algebra background may simply view **W** matrix as weighted average environmental values matrix. Again, each row corresponds to a species.
- This is a Fundamental Matrix of CCA.
- The CCA analyzes species-environmental variations present in this matrix.
- Also, because of how we construct **W**, the solutions to CCA satisfy certain scaling conditions. Two CCA programs may differ in how **W** is prepared. The interpretations may vary depending upon how this matrix is constructed.

More on **W**

$$
WW^t = \begin{pmatrix}
0.014 & -0.02 & 0.015 & 0.019 & 0.02 & 0.011 & 0.003 & 0.001 & -0.007 & -0.025 & -0.025 & -0.026 \\
-0.02 & 0.147 & 0.013 & -0.026 & -0.041 & -0.012 & 0.064 & 0.014 & -0.062 & -0.047 & -0.038 & -0.017 \\
0.015 & 0.013 & 0.028 & 0.019 & 0.017 & 0.012 & 0.022 & 0.001 & -0.027 & -0.048 & -0.045 & -0.042 \\
0.019 & -0.026 & 0.019 & 0.035 & 0.036 & 0.022 & 0.006 & 0.006 & -0.018 & -0.042 & -0.048 & -0.043 \\
0.02 & -0.041 & 0.017 & 0.036 & 0.049 & 0.027 & 0. & 0.007 & 0. & -0.039 & -0.068 & -0.064 \\
0.011 & -0.012 & 0.012 & 0.022 & 0.027 & 0.021 & 0.004 & 0.006 & -0.012 & -0.026 & -0.043 & -0.04 \\
0.003 & 0.064 & 0.022 & 0.006 & 0. & 0.004 & 0.042 & 0.012 & -0.038 & -0.057 & -0.058 & -0.044 \\
0.001 & 0.014 & 0.001 & 0.006 & 0.007 & 0.006 & 0.012 & 0.014 & -0.005 & -0.021 & -0.036 & -0.028 \\
-0.007 & -0.062 & -0.027 & -0.018 & 0. & -0.012 & -0.038 & -0.005 & 0.07 & 0.058 & 0.036 & 0.021 \\
-0.025 & -0.047 & -0.048 & -0.042 & -0.039 & -0.026 & -0.057 & -0.021 & 0.058 & 0.113 & 0.125 & 0.105 \\
-0.025 & -0.038 & -0.045 & -0.048 & -0.068 & -0.043 & -0.058 & -0.036 & 0.036 & 0.125 & 0.189 & 0.164 \\
-0.026 & -0.017 & -0.042 & -0.043 & -0.064 & -0.04 & -0.044 & -0.028 & 0.021 & 0.105 & 0.164 & 0.149
\end{pmatrix}
$$

- We have not shown this matrix in our program as we felt that it may not be of interest to all users. In SAS(IML), you may use W*T(W) to get this matrix.
- The total variation in **W** is the sum of all the diagonal elements of this matrix. Check that it is about 0.87.
- Later you will see that this sum also equals to the sum of all the eigenvalues.
- This matrix may be called inter-species space.
- A more serious researcher may be able to study inter-species distances and correlations using this matrix.
- Note that there are 12 dimensions in this data set. One objective of CCA is to find out what percent of the variations in the 12 dimensions is distributed in the first two dimensions, the first two eigenvectors of this matrix or the left singular vectors of **W** corresponding to the first two largest eigenvalues. The larger the percent, the better the quality of a biplot.

Some More on **W**

$$
\mathbf{W}'\mathbf{W} = \begin{pmatrix} 0.24 & -0.153 & -0.081 & -0.103 & 0.022 & 0.134 \\ -0.153 & 0.161 & 0.062 & 0.09 & -0.08 & -0.073 \\ -0.081 & 0.062 & 0.117 & 0.093 & -0.068 & -0.015 \\ -0.103 & 0.09 & 0.093 & 0.11 & -0.077 & -0.023 \\ 0.022 & -0.08 & -0.068 & -0.077 & 0.122 & -0.039 \\ 0.134 & -0.073 & -0.015 & -0.023 & -0.039 & 0.121 \end{pmatrix}
$$

- We have not shown this matrix in our program as we felt that it may not be of interest to all users. In SAS(IML), you may use T(W)*W to get this matrix.
- Again, the total variation in **W** is the sum of all the diagonal elements of this matrix. Check that it is also about 0.87.
- This matrix may be called inter-environ space.
- A more serious researcher may be able to study inter-environ distances and correlations using this matrix. **A large diagonal value in this matrix also means a long arrow in a biplot.**
- Note that there are 6 dimensions in this data set. One objective of CCA is to find out what percent of the variations in the 6 dimensions is distributed in the first two dimensions, the first two eigenvectors of this matrix or the right singular vectors of **W** corresponding to the first two largest eigenvalues.

P: Left Singular Vectors of **W** or Left Eigenvectors of $\mathbf{WW}^t$

P =

-0.091	-0.14	0.221	-0.251	0.022	0.21
-0.161	0.759	-0.218	0.108	-0.202	-0.121
-0.174	0.016	0.285	-0.469	-0.352	0.231
-0.166	-0.204	0.357	0.178	0.242	-0.114
-0.197	-0.338	0.078	0.196	-0.18	-0.571
-0.134	-0.137	0.106	0.451	-0.325	0.403
-0.209	0.275	0.001	-0.2	0.179	-0.093
-0.101	0.019	-0.168	0.274	0.601	0.458
0.181	-0.319	-0.601	-0.411	0.114	-0.063
0.434	-0.09	-0.291	0.33	-0.375	0.084
0.575	0.1	0.322	-0.166	-0.064	0.222
0.496	0.185	0.317	0.109	0.295	-0.33

M=

-0.684	-1.048	1.652	-1.875	0.161	1.572
-0.662	3.123	-0.899	0.443	-0.833	-0.498
-0.626	0.057	1.026	-1.692	-1.269	0.833
-0.528	-0.649	1.137	0.569	0.773	-0.362
-0.558	-0.955	0.22	0.553	-0.508	-1.613
-0.552	-0.565	0.437	1.855	-1.339	1.657
-0.44	0.579	0.002	-0.42	0.378	-0.196
-0.381	0.071	-0.635	1.036	2.274	1.733
0.513	-0.901	-1.698	-1.161	0.322	-0.179
1.582	-0.328	-1.063	1.205	-1.369	0.306
2.665	0.466	1.491	-0.77	-0.298	1.029
3.384	1.262	2.162	0.743	2.014	-2.249

D=

0.734	0.	0.	0.	0.	0.
0.	0.474	0.	0.	0.	0.
0.	0.	0.27	0.	0.	0.
0.	0.	0.	0.141	0.	0.
0.	0.	0.	0.	0.106	0.
0.	0.	0.	0.	0.	0.06

EigenVals=

0.539	0.	0.	0.	0.	0.
0.	0.224	0.	0.	0.	0.
0.	0.	0.073	0.	0.	0.
0.	0.	0.	0.02	0.	0.
0.	0.	0.	0.	0.011	0.
0.	0.	0.	0.	0.	0.004

- First of all, note that our program cuts off all the zero eigenvalues and the corresponding eigenvectors, called null vectors. **This will not affect our analysis**.
- The first column of **P** provides optimum weights to combine the rows of **W** or the columns of $\mathbf{WW}^t$. That optimum combination of the rows of **W** has the maximum variance, the first eigenvalue (0.54 of 0.87 = 62%) or the the square of the first singular value, the first element of **D** matrix.
- A serious researcher will pay attention to the signs and the magnitude of these elements in **P**. It tells how different rows of **W** (a row vector of size 6, species vectors) be combined in order to explain the maximum possible variance in the first axis.
- In other words, this eigen analysis detects the direction of maximum variance that is present in the inter-species space ($\mathbf{WW}^t$).
- Once the first axis is detected, then the second axis detects the direction of maximum variance that is independent to the first axis. In our case, the second axis (the second column of **P**) detects about 26 percent of the total variance (.224/0.87).
- This data set (**W** matrix) has a good approximation in the first two axes given CCA assumptions.
- The **M** is called generalized singular vector of **A** matrix. It is the same as the **P** matrix adjusted for the species weights.
- The first two columns of the **M** matrix contain the basic coordinates for a species plot in an ordinary xy-coordinate system. These coordinates tells us the relative positions of different species in the overall species cloud which could not have been guessed without this analysis.
- If we use the first two values in **D** matrix to scale the first two columns of **M** matrix, approximaton of inter-species distances and angles improves and at the same time it affects the inter-environ space approximation.

Right Singular Vectors

$$Q=\begin{pmatrix} -0.615 & -0.353 & 0.101 & -0.467 & 0.487 & 0.176 \\ 0.494 & -0.05 & 0.606 & -0.241 & 0.024 & 0.572 \\ 0.319 & -0.333 & -0.657 & -0.505 & -0.255 & 0.189 \\ 0.374 & -0.297 & -0.26 & 0.445 & 0.707 & 0.084 \\ -0.22 & 0.615 & -0.35 & 0.104 & 0.103 & 0.655 \\ -0.296 & -0.543 & 0.026 & 0.51 & -0.432 & 0.412 \end{pmatrix}$$

$$N=\begin{pmatrix} -0.937 & -0.079 & 0.142 & -0.162 & 0.242 & 0.106 \\ 0.748 & 0.057 & 0.46 & -0.176 & 0.017 & 0.44 \\ 0.673 & -0.317 & -0.579 & -0.262 & -0.2 & 0.053 \\ 0.633 & -0.57 & -0.195 & 0.389 & 0.289 & -0.039 \\ -0.399 & 0.79 & -0.094 & -0.256 & 0.092 & 0.368 \\ -0.335 & -0.775 & 0.062 & 0.486 & -0.194 & 0.096 \end{pmatrix}$$

$$D=\begin{pmatrix} 0.734 & 0. & 0. & 0. & 0. & 0. \\ 0. & 0.474 & 0. & 0. & 0. & 0. \\ 0. & 0. & 0.27 & 0. & 0. & 0. \\ 0. & 0. & 0. & 0.141 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0.106 & 0. \\ 0. & 0. & 0. & 0. & 0. & 0.06 \end{pmatrix}$$

$$\text{EigenVals}=\begin{pmatrix} 0.539 & 0. & 0. & 0. & 0. & 0. \\ 0. & 0.224 & 0. & 0. & 0. & 0. \\ 0. & 0. & 0.073 & 0. & 0. & 0. \\ 0. & 0. & 0. & 0.02 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0.011 & 0. \\ 0. & 0. & 0. & 0. & 0. & 0.004 \end{pmatrix}$$

- Note that singular values and eigenvalues are the same for the left vectors and the right vectors.
- The first column of **Q** provides optimum weights to combine the columns of **W** or the columns of $\mathbf{W}^t\mathbf{W}$. That optimum combination of the columns of **W** has the maximum variance, the first eigenvalue (0.53 of 0.87 = 61%) or the the square of the first singular value, the first element of **D** matrix.
- A serious researcher will pay attention to the signs and the magnitude of these elements in **Q**. It tells how different environmental info in the columns of **W** be combined in order to explain the maximum possible variance in the first axis.
- In other words, this eigen analysis detects the direction of maximum variance that is present in the inter-environ space ($\mathbf{W}^t\mathbf{W}$).
- The **N** is called generalized right singular vector of **A** matrix. It is the same as the **Q** matrix adjusted for the square root of the correlation matrix of **Z**.
- The first two columns of the **N** matrix contain the basic coordinates for a environmental plot in an ordinary xy-coordinate system. Generally an arrow is drawn from the origin to these points.These coordinates tells us the relative positions of different environs in the overall environ cloud which could not have been guessed without this analysis.
- If we use the first two values in **D** matrix to scale the first two columns of **N** matrix, approximaton of inter-environ distances and angles improves and at the same time it affects the inter-species space approximation.

Species Scores and Canonical Coefficients

$$\mathbf{U} = \begin{pmatrix} -0.502 & -0.497 & 0.446 & -0.264 & 0.017 & 0.095 \\ -0.486 & 1.479 & -0.243 & 0.062 & -0.089 & -0.03 \\ -0.46 & 0.027 & 0.277 & -0.238 & -0.135 & 0.05 \\ -0.388 & -0.307 & 0.307 & 0.08 & 0.082 & -0.022 \\ -0.409 & -0.452 & 0.059 & 0.078 & -0.054 & -0.097 \\ -0.405 & -0.268 & 0.118 & 0.261 & -0.142 & 0.1 \\ -0.323 & 0.274 & 0. & -0.059 & 0.04 & -0.012 \\ -0.28 & 0.033 & -0.171 & 0.146 & 0.242 & 0.105 \\ 0.376 & -0.427 & -0.458 & -0.163 & 0.034 & -0.011 \\ 1.162 & -0.156 & -0.287 & 0.17 & -0.146 & 0.018 \\ 1.957 & 0.221 & 0.402 & -0.108 & -0.032 & 0.062 \\ 2.484 & 0.598 & 0.583 & 0.105 & 0.214 & -0.136 \end{pmatrix}$$

$$\text{Bmat}= \begin{pmatrix} -0.448 & -0.837 & 0.129 & -1.246 & 1.127 & 0.196 \\ 0.313 & -0.264 & 0.774 & -0.448 & 0.025 & 0.786 \\ 0.049 & -0.538 & -0.782 & -0.952 & -0.355 & 0.436 \\ 0.247 & -0.051 & -0.515 & 0.711 & 1.564 & 0.35 \\ -0.152 & 0.697 & -0.869 & 0.886 & 0.12 & 1.423 \\ -0.287 & -0.257 & -0.143 & 0.863 & -0.996 & 1.052 \end{pmatrix}$$

Linear Combinations of Environmental Variables and Sample Scores

$$\mathbf{X} = \begin{pmatrix}
-1.12 & 1.5 & -0.455 & -0.114 & 0.375 & 0.765 \\
-0.488 & 2.107 & -0.732 & -1.044 & -0.113 & -0.45 \\
-0.471 & 2.343 & -0.047 & -1.565 & 1.341 & -1.489 \\
-0.471 & 2.343 & -0.047 & -1.565 & 1.341 & -1.489 \\
-1.034 & 1.482 & -0.634 & 0.134 & 0.92 & 0.887 \\
-0.876 & 1.987 & -0.422 & 0.935 & 0.602 & 0.665 \\
-0.939 & -0.023 & 0.186 & 0.331 & -0.86 & 0.211 \\
-0.694 & 2.193 & -0.922 & 1.354 & -1.444 & 1.095 \\
-0.521 & 2.467 & -0.209 & -0.407 & 0.125 & -0.742 \\
-0.666 & -0.314 & 1.118 & 0.682 & 0.361 & -1.094 \\
-0.491 & -0.226 & 0.404 & -2.412 & -0.513 & 2.396 \\
-0.858 & -0.982 & 0.609 & -1.335 & -1.748 & -0.874 \\
-0.151 & -0.583 & -0.47 & 0.302 & -0.564 & -0.394 \\
-0.65 & -0.494 & 0.856 & 0.363 & 0.242 & -0.948 \\
-0.434 & -0.535 & 0.651 & 0.652 & 1.901 & 1.511 \\
0.021 & -0.619 & -0.829 & 0.796 & 0.525 & -0.15 \\
0.228 & -0.655 & -1.402 & 0.642 & -0.15 & 0.066 \\
0.512 & -0.697 & -1.579 & -1.255 & 0.235 & -0.997 \\
0.183 & -1.059 & -1.558 & -0.497 & 0.557 & -0.123 \\
1.169 & 0.356 & 2.429 & 0.947 & -1.607 & 0.468 \\
0.659 & -0.49 & -1.864 & -0.187 & -0.047 & -0.54 \\
1.788 & -0.721 & 1.091 & -1.568 & 1.045 & 0.7 \\
2.271 & 0.396 & 0.282 & 0.552 & -1.22 & 0.659 \\
2.256 & 0.297 & -0.449 & 0.648 & -0.804 & 0.632 \\
2.787 & 0.493 & 0.405 & -2.274 & 1.434 & -1.587 \\
2.036 & 0.594 & -0.298 & 0.888 & -1.239 & 0.07 \\
2.138 & -0.351 & -0.002 & -0.823 & 0.468 & 0.668 \\
2.974 & 1.302 & 1.953 & 0.389 & 1.376 & -0.86
\end{pmatrix}$$

$$\mathbf{X}^{*} = \begin{pmatrix}
-0.381 & 0.545 & -0.026 & -0.051 & -0.014 & -0.009 \\
-0.388 & 0.529 & -0.031 & -0.006 & 0.008 & 0. \\
-0.388 & 0.622 & -0.069 & -0.016 & 0. & 0.002 \\
-0.38 & 0.374 & 0.019 & -0.038 & 0.007 & 0.006 \\
-0.373 & 0.434 & 0.006 & -0.069 & -0.01 & -0.005 \\
-0.359 & 0.546 & -0.082 & -0.003 & 0.033 & -0.002 \\
-0.357 & -0.058 & 0.065 & 0.031 & -0.024 & -0.001 \\
-0.359 & 0.482 & -0.085 & 0.041 & -0.002 & 0.008 \\
-0.334 & 0.516 & -0.105 & -0.014 & -0.009 & -0.01 \\
-0.324 & -0.099 & 0.085 & 0.025 & 0.012 & -0.002 \\
-0.313 & -0.134 & 0.086 & -0.037 & 0.005 & 0.008 \\
-0.321 & -0.083 & 0.051 & -0.027 & -0.015 & -0.011 \\
-0.31 & -0.103 & 0.066 & 0.009 & -0.016 & -0.004 \\
-0.278 & -0.161 & 0.065 & -0.002 & 0. & -0.003 \\
-0.226 & -0.103 & 0.036 & 0.018 & 0.016 & 0.006 \\
-0.082 & -0.102 & -0.027 & 0.022 & 0.01 & 0.004 \\
0. & -0.167 & -0.088 & 0.014 & -0.003 & -0.003 \\
0.164 & -0.274 & -0.238 & -0.04 & 0.034 & -0.006 \\
0.228 & -0.216 & -0.244 & -0.014 & 0.025 & 0.002 \\
0.316 & 0.035 & -0.008 & -0.02 & -0.018 & 0.02 \\
0.394 & -0.235 & -0.254 & -0.039 & -0.002 & 0.006 \\
0.614 & -0.195 & -0.245 & -0.093 & 0. & 0.004 \\
1.227 & 0.021 & 0.001 & -0.01 & -0.021 & 0.007 \\
1.187 & 0.164 & 0.005 & 0.018 & -0.018 & -0.006 \\
1.41 & 0.113 & 0.129 & -0.024 & -0.014 & 0.014 \\
1.548 & 0.059 & 0.044 & 0.052 & -0.033 & 0. \\
1.666 & 0.118 & 0.146 & 0.012 & -0.01 & 0.002 \\
1.836 & 0.369 & 0.357 & 0.039 & 0.086 & -0.051
\end{pmatrix}$$

Correlation Structures

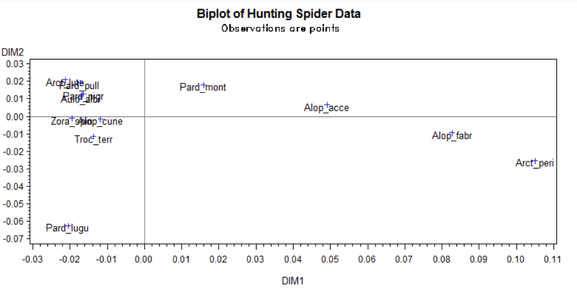
$$Corr(\mathbf{Z},\mathbf{X}) = \begin{pmatrix} -0.937 & 0.748 & 0.673 & 0.633 & -0.399 & -0.335 \\ -0.079 & 0.057 & -0.317 & -0.57 & 0.79 & -0.775 \\ 0.142 & 0.46 & -0.579 & -0.195 & -0.094 & 0.062 \\ -0.162 & -0.176 & -0.262 & 0.389 & -0.256 & 0.486 \\ 0.242 & 0.017 & -0.2 & 0.289 & 0.092 & -0.194 \\ 0.106 & 0.44 & 0.053 & -0.039 & 0.368 & 0.096 \end{pmatrix}$$

$$Corr(\mathbf{X},\mathbf{X}^*) = \begin{pmatrix} 0.958 & 0. & 0. & 0. & 0. & 0. \\ 0. & 0.948 & 0. & 0. & 0. & 0. \\ 0. & 0. & 0.665 & 0. & 0. & 0. \\ 0. & 0. & 0. & 0.711 & 0. & 0. \\ 0. & 0. & 0. & 0. & 0.624 & 0. \\ 0. & 0. & 0. & 0. & 0. & 0.399 \end{pmatrix}$$

$$Corr(\mathbf{Z},\mathbf{X}^*) = \begin{pmatrix} -0.898 & -0.075 & 0.094 & -0.115 & 0.151 & 0.042 \\ 0.717 & 0.054 & 0.306 & -0.125 & 0.01 & 0.176 \\ 0.645 & -0.301 & -0.385 & -0.186 & -0.125 & 0.021 \\ 0.607 & -0.54 & -0.129 & 0.277 & 0.18 & -0.016 \\ -0.382 & 0.748 & -0.062 & -0.182 & 0.057 & 0.147 \\ -0.321 & -0.735 & 0.041 & 0.346 & -0.121 & 0.038 \end{pmatrix}$$

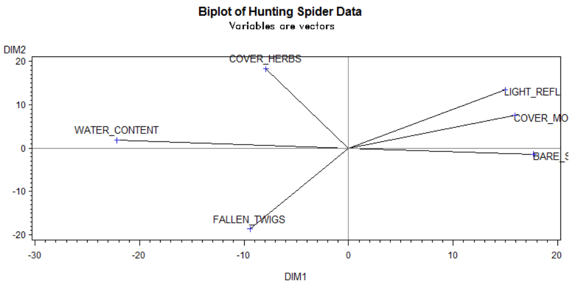
A two dimensional plot of Species Coordinates

$$
SpeCoord = \begin{pmatrix}
-0.501925 & -0.496608 \\
-0.486002 & 1.47922 \\
-0.459762 & 0.0271519 \\
-0.387949 & -0.307493 \\
-0.409366 & -0.452335 \\
-0.405182 & -0.26772 \\
-0.323372 & 0.274468 \\
-0.279575 & 0.0334517 \\
0.376497 & -0.426718 \\
1.16178 & -0.15555 \\
1.95676 & 0.220531 \\
2.4845 & 0.598033
\end{pmatrix}
\qquad
Hill = \begin{pmatrix}
-0.0212292 & -0.0210043 \\
-0.0205557 & 0.0625645 \\
-0.0194459 & 0.0011484 \\
-0.0164085 & -0.0130056 \\
-0.0173143 & -0.0191317 \\
-0.0171374 & -0.0113233 \\
-0.0136772 & 0.0116088 \\
-0.0118248 & 0.00141486 \\
0.0159241 & -0.0180483 \\
0.049138 & -0.00657907 \\
0.0827622 & 0.00932747 \\
0.105083 & 0.0252941
\end{pmatrix}
$$



A two dimensional plot of Env Coordinates

$$EnvCoord = \begin{pmatrix} -0.936753 & -0.0790415 \\ 0.748498 & 0.0572778 \\ 0.672746 & -0.317131 \\ 0.6332 & -0.569844 \\ -0.398647 & 0.789517 \\ -0.334521 & -0.775032 \end{pmatrix} \qquad Hill = \begin{pmatrix} -22.1478 & -1.86879 \\ 17.6969 & 1.35423 \\ 15.9058 & -7.49797 \\ 14.9709 & -13.4729 \\ -9.42527 & 18.6667 \\ -7.90915 & -18.3242 \end{pmatrix}$$



Together

$$Biplot1 = \begin{pmatrix} -0.501925 & -0.496608 \\ -0.486002 & 1.47922 \\ -0.459762 & 0.0271519 \\ -0.387949 & -0.307493 \\ -0.409366 & -0.452335 \\ -0.405182 & -0.26772 \\ -0.323372 & 0.274468 \\ -0.279575 & 0.0334517 \\ 0.376497 & -0.426718 \\ 1.16178 & -0.15555 \\ 1.95676 & 0.220531 \\ 2.4845 & 0.598033 \\ -0.936753 & -0.0790415 \\ 0.748498 & 0.0572778 \\ 0.672746 & -0.317131 \\ 0.6332 & -0.569844 \\ -0.398647 & 0.789517 \\ -0.334521 & -0.775032 \end{pmatrix}$$

$$Biplot2 = \begin{pmatrix} -0.0212292 & -0.0210043 \\ -0.0205557 & 0.0625645 \\ -0.0194459 & 0.0011484 \\ -0.0164085 & -0.0130056 \\ -0.0173143 & -0.0191317 \\ -0.0171374 & -0.0113233 \\ -0.0136772 & 0.0116088 \\ -0.0118248 & 0.00141486 \\ 0.0159241 & -0.0180483 \\ 0.049138 & -0.00657907 \\ 0.0827622 & 0.00932747 \\ 0.105083 & 0.0252941 \\ -22.1478 & -1.86879 \\ 17.6969 & 1.35423 \\ 15.9058 & -7.49797 \\ 14.9709 & -13.4729 \\ -9.42527 & 18.6667 \\ -7.90915 & -18.3242 \end{pmatrix}$$

$$biplot3 = \begin{pmatrix} -0.0212292 & -0.0210043 \\ -0.0205557 & 0.0625645 \\ -0.0194459 & 0.0011484 \\ -0.0164085 & -0.0130056 \\ -0.0173143 & -0.0191317 \\ -0.0171374 & -0.0113233 \\ -0.0136772 & 0.0116088 \\ -0.0118248 & 0.00141486 \\ 0.0159241 & -0.0180483 \\ 0.049138 & -0.00657907 \\ 0.0827622 & 0.00932747 \\ 0.105083 & 0.0252941 \\ -0.0442956 & -0.00373759 \\ 0.0353938 & 0.00270846 \\ 0.0318117 & -0.0149959 \\ 0.0299417 & -0.0269459 \\ -0.0188505 & 0.0373334 \\ -0.0158183 & -0.0366485 \end{pmatrix}$$