Step 5: Conduct Analysis

Model Parameterization:

- Dropped species with fewer than 5 occurrences
- Log-transformed species abundances
- Row-normalized species’ log abundances (chord distance)
- Selected smallest “best” subset of variables from each set of independent variables
- Chose RDA (in combo with chord distance)

The CCA Algorithm

Assign arbitrary weights to samples

→ Compute species scores as weighted averages

→ Compute sample scores as weighted averages

→ Regress sample scores on environmental variables and assign new sample scores as the predicted values from regression

→ Rescale (standardize) sample scores

→ Subsequent axes computed from residuals

CCA estimates species optima, regression coefficients, and site scores using a Gaussian weighted-averaging model combined with regression.

- Gradients are assumed to be linear combinations of the environmental variables.
- Key assumption is that of unimodal species distributions along environmental gradients.
Step 6: Interpret Results

- **Effectiveness of the Constrained Ordination**
  - Total inertia, eigenvalues, percent variance explained, significance tests, species-environment correlation, cumulative species/sample fit.

- **Interpreting the Constrained Ordination**
  - Canonical coefficients, inter-set and intra-set correlations, biplot scores.
  - The Triplot: sample vs species scaling, centroid rule vs biplot rule, general interpretation.
  - Interpreting species-environment relationships.

---

**Effectiveness of the Constrained Ordination**

**Total inertia**
- Sum of the eigenvalues; total “variance” in the species data, some of which is accounted for by the environmental constraints.

**Eigenvalues**
- Variance in the community matrix attributed to a particular axis; measure of importance of the ordination axis.

---

Call: rda(formula = y.chord ~ BATOTAL + MGF + CLS + MLS + CCF + MCP + MLS, data = x.plot)

Partitioning of variance:

<table>
<thead>
<tr>
<th></th>
<th>Inertia</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3.326</td>
<td>1.00</td>
</tr>
<tr>
<td>Constrained</td>
<td>0.0731</td>
<td>0.22</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>0.2595</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Eigenvalues, and their contribution to the variance

<table>
<thead>
<tr>
<th></th>
<th>Lambda</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RDA1</td>
<td>RDA2</td>
<td>RDA3</td>
<td>RDA4</td>
<td>RDA5</td>
<td>RDA6</td>
<td>RDA7</td>
<td>PC1</td>
<td>PC2</td>
</tr>
<tr>
<td>lambda</td>
<td>0.0503</td>
<td>0.0115</td>
<td>0.0058</td>
<td>0.0030</td>
<td>0.0003</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0013</td>
<td>0.0021</td>
<td>0.0022</td>
</tr>
<tr>
<td>accounted</td>
<td>0.1527</td>
<td>0.1874</td>
<td>0.2039</td>
<td>0.2130</td>
<td>0.2170</td>
<td>0.2189</td>
<td>0.2189</td>
<td>0.2198</td>
<td>0.2050</td>
<td>0.1424</td>
</tr>
</tbody>
</table>
Effectiveness of the Constrained Ordination

Percent “variance” of species data

- Proportion of species variance (total inertia) explained by each axis. Computed as eigenvalue/total inertia (reported as cumulative %).
- For ecological data the percentage of explained variance is usually low; often ~ 10%. Not to worry, as it is an inherent feature of data with a strong presence/absence aspect.

```
Call:
  rda(formula = y.chord ~ BATOTAL + MGF + CLS + MLS + CCP + MCP + MLS, data = x.plot)

Partitioning of variance:

                   Inertia Proportion
Total             0.3326     1.00
Constrained       0.0731     0.22
Unconstrained     0.2595     0.78

Eigenvalues, and their contribution to the variance

       RDA1  RDA2  RDA3  RDA4  RDA5  RDA6  RDA7  SC1  SC2  SC3
lambda 0.0508 0.0115 0.00551 0.00302 0.00134 0.000606 0.000311 0.0271 0.0203 0.0167
accounted 0.1527 0.1874 0.20397 0.21303 0.21707 0.218891 0.219827 0.0815 0.1424 0.1926
```

Effectiveness of the Constrained Ordination

Significance Tests

- Monte Carlo global permutation tests of significance of canonical axes.

```
Model: rda(formula = y.chord ~ BATOTAL + MGF + C
            Df  Var Pr(>F)    
RDA1 1 0.05 13.8962 200.00 < 0.005 ***
RDA2 1 0.01 3.1591 200.00 < 0.005 ***
RDA3 1 0.01 1.5067 200.00 < 0.005 ***
RDA4 1 0.0030159 0.8252 200.00 < 0.005 ***
RDA5 1 0.0013428 0.3674 200.00 < 0.005 ***
RDA6 1 0.00006057 0.1657 400.00 0.0225 *
RDA7 1 0.0003113 0.0852 100.00 0.1700
Residual 71 0.26
```

- Test based on 1st canonical axis has maximum power against the alternative hypothesis that there is a single dominating gradient that determines the relation between the species and environment.
Scores for Sites and Species

- Ordination scores (coordinates on ordination axes) given for each site (sample) and each species. Two sets of scores are produced for sites (samples):
  - *Linear combination (LC) scores* – linear combinations of the independent variables. LC scores are in environmental space, each axis formed by linear combinations of environmental variables.
  - *Weighted averages (WA) scores* – weighted averages (sums) of the species scores that are as similar to LC scores as possible. WA scores are in species space.

  LC scores show were the site *should* be; the WA scores show where the site *is*. 

---

**Effectiveness of the Constrained Ordination**

LC scores (samples)  
WA scores (samples)
Effectiveness of the Constrained Ordination

LC versus WA Scores

- Longstanding debate (confusion) over whether to use LC or WA scores.
- LC scores are very sensitive to noise in the environmental data, whereas the WA scores are much more stable, being largely determined by the species data rather than being a direct projection in environmental space.

“Use of LC scores will misrepresent the observed community relationships unless the environmental data are both noiseless and meaningful.... CCA (RDA) with LC scores is inappropriate where the objective is to describe community structure.” (McCune)

Effectiveness of the Constrained Ordination

Species-environmental correlations

- The multiple correlation coefficient of the final regression. Correlation between site scores computed as weighted averages of species scores (WA’s) and site scores computed as linear combinations of environmental variables (LC’s); measures how well the extracted variation in community composition can be explained by environmental variables.

```
RDA1  RDA2  RDA3  RDA4  RDA5  RDA6  RDA7
0.0467644 0.7043517 0.5683632 0.4764235 0.3371492 0.2756923 0.2061113
```

“This is a bad measure of goodness of ordination, because it is sensitive to extreme scores (like correlations are), and very sensitive to overfitting or using too many constraints.” (Oksanen)
Effectiveness of the Constrained Ordination

Goodness-of-Fit for Species/Samples

- Ordination diagnostic to find out which species/samples are ill-represented in the ordination.
  - % variance (inertia) explained by selected axes.

Plot species with >10% explained

<table>
<thead>
<tr>
<th>AMCR</th>
<th>AMGO</th>
<th>AMRO</th>
<th>BCCR</th>
<th>BEKI</th>
<th>BBR</th>
<th>BOHA</th>
<th>BHBO</th>
<th>BRBR</th>
<th>BCR</th>
<th>BTPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.015</td>
<td>0.220</td>
<td>0.014</td>
<td>0.025</td>
<td>0.003</td>
<td>0.156</td>
<td>0.090</td>
<td>0.003</td>
<td>0.054</td>
<td>0.467</td>
<td>0.102</td>
</tr>
<tr>
<td>BDRR</td>
<td>BSSH</td>
<td>CBCH</td>
<td>CBCH</td>
<td>CERN</td>
<td>CONI</td>
<td>COPA</td>
<td>COYE</td>
<td>DENT</td>
<td>DING</td>
<td>DORW</td>
</tr>
<tr>
<td>0.005</td>
<td>0.032</td>
<td>0.144</td>
<td>0.139</td>
<td>0.011</td>
<td>0.009</td>
<td>0.010</td>
<td>0.076</td>
<td>0.003</td>
<td>0.083</td>
<td>0.003</td>
</tr>
<tr>
<td>GCKI</td>
<td>GROW</td>
<td>GROU</td>
<td>GALC</td>
<td>GENU</td>
<td>HEBR</td>
<td>HEBR</td>
<td>HEMU</td>
<td>HOMR</td>
<td>HUIU</td>
<td>HULJ</td>
</tr>
<tr>
<td>0.159</td>
<td>0.003</td>
<td>0.076</td>
<td>0.256</td>
<td>0.106</td>
<td>0.012</td>
<td>0.159</td>
<td>0.493</td>
<td>0.210</td>
<td>0.001</td>
<td>0.009</td>
</tr>
<tr>
<td>MOKA</td>
<td>MQGQ</td>
<td>NOPL</td>
<td>ONCI</td>
<td>OBFL</td>
<td>PISI</td>
<td>PICO</td>
<td>PISI</td>
<td>PITU</td>
<td>PITU</td>
<td>POPQ</td>
</tr>
<tr>
<td>0.260</td>
<td>0.031</td>
<td>0.174</td>
<td>0.490</td>
<td>0.092</td>
<td>0.003</td>
<td>0.006</td>
<td>0.284</td>
<td>0.117</td>
<td>0.007</td>
<td>0.060</td>
</tr>
<tr>
<td>RBSA</td>
<td>RBOE</td>
<td>BSTO</td>
<td>RTIA</td>
<td>RUGA</td>
<td>RUGU</td>
<td>SOWU</td>
<td>SOWU</td>
<td>SOWU</td>
<td>SOWU</td>
<td>SOWU</td>
</tr>
<tr>
<td>0.008</td>
<td>0.002</td>
<td>0.292</td>
<td>0.063</td>
<td>0.021</td>
<td>0.243</td>
<td>0.046</td>
<td>0.412</td>
<td>0.005</td>
<td>0.005</td>
<td>0.019</td>
</tr>
<tr>
<td>SMTH</td>
<td>TGSO</td>
<td>TSHM</td>
<td>TUUV</td>
<td>VATE</td>
<td>VGSW</td>
<td>WAVI</td>
<td>WELR</td>
<td>WELR</td>
<td>WELR</td>
<td>WFL</td>
</tr>
<tr>
<td>0.198</td>
<td>0.053</td>
<td>0.040</td>
<td>0.010</td>
<td>0.304</td>
<td>0.157</td>
<td>0.143</td>
<td>0.418</td>
<td>0.134</td>
<td>0.055</td>
<td>0.177</td>
</tr>
<tr>
<td>WIRA</td>
<td>WIRE</td>
<td>WRBD</td>
<td>WAPC</td>
<td>TORA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.039</td>
<td>0.439</td>
<td>0.146</td>
<td>0.071</td>
<td>0.038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpreting the Constrained Ordination

Canonical Coefficients

- Regression coefficients of the linear combinations of environmental variables that define the ordination axes; they are eigenvector coefficients and not very useful for interpretation.

Unstandardized coefficients

<table>
<thead>
<tr>
<th>RDA1</th>
<th>RDA2</th>
<th>RDA3</th>
<th>RDA4</th>
<th>RDA5</th>
<th>RDA6</th>
<th>RDA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATOTAL</td>
<td>0.00017</td>
<td>-0.00001</td>
<td>-0.00023</td>
<td>-0.00032</td>
<td>0.00179</td>
<td>-0.00050</td>
</tr>
<tr>
<td>MCF</td>
<td>-0.00013</td>
<td>0.00054</td>
<td>-0.00001</td>
<td>-0.00144</td>
<td>-0.00020</td>
<td>-0.00017</td>
</tr>
<tr>
<td>CLSS</td>
<td>0.00074</td>
<td>0.00019</td>
<td>-0.00026</td>
<td>-0.00023</td>
<td>-0.00075</td>
<td>0.00036</td>
</tr>
<tr>
<td>LCS</td>
<td>0.00070</td>
<td>0.00003</td>
<td>0.00042</td>
<td>-0.00027</td>
<td>-0.00066</td>
<td>-0.00039</td>
</tr>
<tr>
<td>CCP</td>
<td>0.00041</td>
<td>-0.00046</td>
<td>-0.00028</td>
<td>-0.00060</td>
<td>-0.00059</td>
<td>-0.00033</td>
</tr>
<tr>
<td>MCFP</td>
<td>0.00044</td>
<td>-0.00056</td>
<td>0.00024</td>
<td>-0.00098</td>
<td>-0.00015</td>
<td>0.00064</td>
</tr>
<tr>
<td>HLS</td>
<td>0.00056</td>
<td>-0.00016</td>
<td>0.00109</td>
<td>-0.00050</td>
<td>-0.00052</td>
<td>0.00033</td>
</tr>
</tbody>
</table>

Standardized coefficients

<table>
<thead>
<tr>
<th>RDA1</th>
<th>RDA2</th>
<th>RDA3</th>
<th>RDA4</th>
<th>RDA5</th>
<th>RDA6</th>
<th>RDA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATOTAL</td>
<td>0.00350</td>
<td>-0.00026</td>
<td>-0.00479</td>
<td>-0.00669</td>
<td>0.03706</td>
<td>-0.01038</td>
</tr>
<tr>
<td>MCF</td>
<td>-0.00244</td>
<td>0.01500</td>
<td>-0.00013</td>
<td>-0.02776</td>
<td>-0.00395</td>
<td>-0.00230</td>
</tr>
<tr>
<td>CLSS</td>
<td>0.02963</td>
<td>0.00780</td>
<td>-0.01050</td>
<td>-0.00926</td>
<td>-0.03020</td>
<td>0.01580</td>
</tr>
<tr>
<td>LCS</td>
<td>0.02550</td>
<td>0.00122</td>
<td>0.01579</td>
<td>-0.01015</td>
<td>-0.02490</td>
<td>-0.00731</td>
</tr>
<tr>
<td>CCP</td>
<td>0.01427</td>
<td>-0.01626</td>
<td>-0.00992</td>
<td>-0.02101</td>
<td>-0.02083</td>
<td>-0.01168</td>
</tr>
<tr>
<td>MCFP</td>
<td>0.01040</td>
<td>-0.03122</td>
<td>0.00563</td>
<td>-0.02095</td>
<td>-0.00362</td>
<td>0.01992</td>
</tr>
<tr>
<td>HLS</td>
<td>0.00960</td>
<td>-0.00246</td>
<td>0.01736</td>
<td>-0.00797</td>
<td>-0.00826</td>
<td>0.00525</td>
</tr>
</tbody>
</table>
Interpreting the Constrained Ordination

**IntRA-set Correlation (structure) Coefficients**

- Correlations between sample scores derived as linear combinations of environmental variables (*L.C. scores*) and environmental variables (i.e., loadings); they are related to the canonical coefficients but differ in several important aspects.

<table>
<thead>
<tr>
<th></th>
<th>RDA1</th>
<th>RDA2</th>
<th>RDA3</th>
<th>RDA4</th>
<th>RDA5</th>
<th>RDA6</th>
<th>RDA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATOTAL</td>
<td>0.683</td>
<td>0.078</td>
<td>-0.135</td>
<td>-0.084</td>
<td>0.678</td>
<td>-0.180</td>
<td>-0.095</td>
</tr>
<tr>
<td>MGF</td>
<td>-0.410</td>
<td>0.632</td>
<td>0.020</td>
<td>-0.646</td>
<td>-0.091</td>
<td>-0.082</td>
<td>0.024</td>
</tr>
<tr>
<td>CLS</td>
<td>0.575</td>
<td>0.401</td>
<td>-0.543</td>
<td>0.160</td>
<td>-0.001</td>
<td>0.391</td>
<td>-0.189</td>
</tr>
<tr>
<td>MLS</td>
<td>0.455</td>
<td>0.131</td>
<td>0.586</td>
<td>0.113</td>
<td>-0.028</td>
<td>-0.455</td>
<td>0.461</td>
</tr>
<tr>
<td>CCP</td>
<td>-0.105</td>
<td>-0.615</td>
<td>-0.420</td>
<td>-0.300</td>
<td>-0.259</td>
<td>-0.487</td>
<td>-0.201</td>
</tr>
<tr>
<td>MGF</td>
<td>-0.076</td>
<td>-0.466</td>
<td>0.169</td>
<td>-0.389</td>
<td>0.176</td>
<td>0.660</td>
<td>0.361</td>
</tr>
<tr>
<td>HLS</td>
<td>0.084</td>
<td>-0.051</td>
<td>0.589</td>
<td>-0.080</td>
<td>0.007</td>
<td>0.107</td>
<td>-0.791</td>
</tr>
</tbody>
</table>

Interpreting the Constrained Ordination

**IntRA-set Correlation (structure) Coefficients**

- Both canonical coefficients and IntRA-set correlations relate to the rate of change in community composition per unit change in the corresponding environmental variable – but for *canonical coefficients* it is assumed that other environmental variables are being held constant, while for *IntRA-set correlations* other environmental variables are assumed to covary as they do in the data set.

- When variables are intercorrelated *canonical coefficients* become unstable – the multicollinearity problem; however, *IntRA-set correlations* are not effected.
Interpreting the Constrained Ordination

**IntRA-set Correlation (structure) Coefficients**

- Canonical coefficients and IntRA-set correlations indicate which environmental variables are more influential in structuring the ordination, but cannot be viewed as an independent measure of the strength of relationship between communities and environmental variables.

- Probably only meaningful if the site scores are derived as LC scores (which is generally not recommended).

> “They have all the problems of correlations, like being sensitive to extreme values, and they focus on the relationship between single constraints and single axes instead of multivariate analysis.” (Oksanen)

---

**IntER-set Correlation (structure) Coefficients**

- Correlations between species-derived sample scores (WA scores) and environmental variables.
  - Should be higher than correlations derived from unconstrained ordination.
  - Generally lower than IntRA-set correlations.
  - IntRA-set correlation x species-environment correlation.

<table>
<thead>
<tr>
<th></th>
<th>RDA1</th>
<th>RDA2</th>
<th>RDA3</th>
<th>RDA4</th>
<th>RDA5</th>
<th>RDA6</th>
<th>RDA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATOTAL</td>
<td>0.579</td>
<td>0.055</td>
<td>-0.077</td>
<td>-0.040</td>
<td>0.229</td>
<td>-0.050</td>
<td>-0.020</td>
</tr>
<tr>
<td>MGF</td>
<td>-0.347</td>
<td>0.445</td>
<td>0.011</td>
<td>-0.308</td>
<td>-0.031</td>
<td>-0.022</td>
<td>0.005</td>
</tr>
<tr>
<td>CLS</td>
<td>0.487</td>
<td>0.282</td>
<td>-0.308</td>
<td>0.076</td>
<td>0.000</td>
<td>0.108</td>
<td>-0.039</td>
</tr>
<tr>
<td>MLS</td>
<td>0.385</td>
<td>0.092</td>
<td>0.333</td>
<td>0.054</td>
<td>-0.010</td>
<td>-0.125</td>
<td>0.095</td>
</tr>
<tr>
<td>CCP</td>
<td>-0.089</td>
<td>-0.433</td>
<td>-0.239</td>
<td>-0.143</td>
<td>-0.087</td>
<td>-0.134</td>
<td>-0.041</td>
</tr>
<tr>
<td>MCP</td>
<td>-0.064</td>
<td>-0.328</td>
<td>0.096</td>
<td>-0.185</td>
<td>0.059</td>
<td>0.182</td>
<td>0.074</td>
</tr>
<tr>
<td>HLS</td>
<td>0.071</td>
<td>-0.036</td>
<td>0.335</td>
<td>-0.038</td>
<td>0.002</td>
<td>0.030</td>
<td>-0.163</td>
</tr>
</tbody>
</table>
**Interpreting the Constrained Ordination**

**IntER-set Correlation (structure) Coefficients**

- Like IntRA-set correlations, they do not become unstable when the environmental variables are strongly correlated with each other (multicollinear).
- Probably only meaningful if the site scores are derived as WA scores (which is generally recommended).

“They have all the problems of correlations, like being sensitive to extreme values, and they focus on the relationship between single constraints and single axes instead of multivariate analysis.” (Oksanen)

---

**Interpreting the Constrained Ordination**

**Biplot Scores for Environmental Variables**

- Environmental variables are typically represented as vectors radiating from the centroid of the ordination.
- Biplot scores give the coordinates of the heads of the environmental vectors, and are based on the IntRA-set correlations.

**Biplot scores for constraining variables**

<table>
<thead>
<tr>
<th></th>
<th>RDA1</th>
<th>RDA2</th>
<th>RDA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATOTAL</td>
<td>0.6833</td>
<td>0.0779</td>
<td>-0.135</td>
</tr>
<tr>
<td>MSF</td>
<td>-0.4096</td>
<td>0.6319</td>
<td>0.020</td>
</tr>
<tr>
<td>CLS</td>
<td>0.5748</td>
<td>0.4011</td>
<td>-0.543</td>
</tr>
<tr>
<td>MLS</td>
<td>0.4546</td>
<td>0.1309</td>
<td>0.586</td>
</tr>
<tr>
<td>CCP</td>
<td>-0.1055</td>
<td>-0.6153</td>
<td>-0.420</td>
</tr>
<tr>
<td>MCP</td>
<td>-0.0760</td>
<td>-0.4662</td>
<td>0.169</td>
</tr>
<tr>
<td>HLS</td>
<td>0.0842</td>
<td>-0.0507</td>
<td>0.589</td>
</tr>
</tbody>
</table>
The Canonical Triplot

- The triplot displays the major patterns in the species data with respect to the environmental variables.

\[ \text{Tri} = (1) \text{ Samples} \quad (2) \text{ Species} \quad (3) \text{ Environment} \]

- But there are a number of ways to scale the triplot which affect the exact quantitative interpretation of the relationships shown.

---

The Canonical Triplot

**Sample Scaling**

- Optimally displays *inter-sample* relationships.

- Variance of *sample* scores on each axis reflects the importance of the axis as measured by the eigenvalue, whereas the variance of the *species* scores along the axes are equal.
Species Scaling

- Optimally displays *inter*-species relationships.

- Variance of *species* scores on each axis reflects the importance of the axis as measured by the eigenvalue, whereas the variance of the *sample* scores along the axes are equal.

*Generally the preferred scaling*

The Centroid Principle

- Way of interpreting ordination diagrams in *unimodal* methods (CCA).

- Each species' point is at the centroid (weighted average) of the sample points where it occurs (i.e., center of its niche); the samples that contain a particular species are scattered around that species' point in the diagram.
The Biplot Rule

- Way of interpreting ordination diagrams in linear methods (RDA).
- An arrow through the species points in direction of maximum change in abundance of the species.
- The order of sites projected onto arrow gives the inferred ranking of the relative abundance of the species across sites.

The Canonical Triplot

General Interpretation – regardless of scaling

- Sample locations indicate their compositional similarity to each other.
- Samples tend to be dominated by the species that are located near them (or projected toward them) in ordination space.
The Canonical Triplot

General Interpretation – regardless of scaling

- *Species* locations indicate their distributional similarity to each other.
- *Species* tend to be most present and abundant in the sites that are located near them (or in the direction of their projection) in ordination space.

---

The Canonical Triplot

General Interpretation – regardless of scaling

- **Length** of environmental vector indicates its importance to the ordination.
- **Direction** of the vector indicates its correlation with each of the axes.
- **Angles** between vectors indicate the correlation between the environmental variables themselves.
The Canonical Triplot

General Interpretation – regardless of scaling

- In unimodal model (CCA), perpendiculars drawn from species to environmental arrow gives approximate ranking of species response to that variable, and whether species has higher-than-average or lower-than average optimum on that environmental variable.

![Diagram showing Canonical Triplot for CCA](image)

Higher-than-average

Lower-than-average

The Canonical Triplot

General Interpretation – regardless of scaling

- In linear model (RDA), direction of arrow drawn from orgin to species and environmental arrow gives approximate relationship between species and environmental variable; i.e., whether species has positive or negative relationship with that environmental variable.

![Diagram showing Canonical Triplot for RDA](image)

Positive correlation

Negative correlation
The Canonical Triplot

General Interpretation – regardless of scaling

- Perpendiculars drawn from *samples* to environmental arrows gives approximate ranking of sample values for that environmental variable.

- And whether a *sample* has a higher-than-average or lower-than-average value on that environmental variable.

Nominal Environmental Variables

- It is more natural to represent nominal environmental variables by points instead of arrows.

- The points are located at the *centroid* of sites that belong to that class.

- Classes containing sites with high values for a species will tend to lie close to that species point.
Don’t worry; focus on relationships

The Canonical Triplot

Samples-scaling  Species-scaling

Don’t worry; focus on relationships

Major Differences Between RDA and CCA

CCA/dCCA

- **Chi-square distance** among samples preserved (approximates unimodal species responses).
- **Rare species** are given disproportionate weight in the analysis.
- **Unequal weighting of sites** in the regression (equal to site totals).

RDA

- **Euclidean distance** among samples preserved (appropriate with linear species responses and can approximate unimodal responses with appropriate data transformations).
- **Rare species** are not given high weight in the analysis.
- **Equal weighting of sites** in the regression.
Major Differences Between RDA and CCA

CCA/dCCA
- Species-environment correlation equals the correlation between the site scores that are weighted averages of the species scores (WA’s) and the site scores that are a linear combination of the environmental variables (LC’s).
- Ordination diagram can be interpreted using the centroid principle.

RDA
- Species-environment correlation equals the correlation between the site scores that are weighted sums of the species scores (WA’s) and the site scores that are a linear combination of the environmental variables (LC’s).
- Ordination diagram can be interpreted using the biplot rule.

Usefulness of Constrained Ordination

Displaying Interpretable Relationships
- Constrained ordination (as with any ordination) is ultimately about portraying meaningful species-environment relationships.
Usefulness of Constrained Ordination

Displaying Interpretable Relationships

• Constrained ordination (as with any ordination) is ultimately about portraying meaningful species-environment relationships.
Multivariate Regression Trees (MRT)

- **Nonparametric** procedure useful for exploration, description, and prediction of species-environment relationships (De’ath 2002).

- **Natural extension of Univariate Regression Trees (URT):** with the univariate response of URT being replaced by a multivariate response.

- **Recursive partitioning** of the data space such that the ‘populations’ within each partition become more and more homogeneous.

Important Characteristics of MRT

- **Makes no assumption** about the form of the relationships between species and their environment.

- Form of **multivariate regression** in that the response is explained, and can be predicted, by the explanatory variables.

- **Method of constrained cluster analysis,** because it determines clusters that are similar in a chosen measure of species dissimilarity (e.g., assemblage type), with each cluster being defined by a set of environmental values (e.g., habitat type).
Univariate Regression Trees (URT)

1. At each node, the tree algorithm searches through the variables one by one, beginning with $x_1$ and continuing up to $x_M$.

2. For each variable it finds the best split (minimizes the total sums of squares or sums of absolute deviations about the median).

3. Then it compares the $M$ best single-variable splits and selects the best of the best.

4. Recursively partition each node until an over-large tree is grown and then pruned back to the desired size (usually based on cross-validated predicted mean square error).

5. Describe each terminal node (mean, median, variance) and the overall fit of the tree (% variance explained).

Multivariate Regression Trees (MRT)

- Redefine node impurity by summing the univariate impurity measure over the multivariate response.
  - Minimize the sums of squared Euclidean distances (SSD) of sites about the node centroid.

- Additional ways to interpret the results of MRT:
  - Identification of species that most strongly determine the splits.
  - Tree biplots to represent group means and species and site information.
  - Identification of species that best characterize the groups (ISA).
Multivariate Regression Trees (MRT)

- Compare the MRT solution to *unconstrained clustering* using a metric equivalent to the MRT impurity and the same number of clusters.

- MRT can be extended to work directly from a *dissimilarity matrix*; clusters can be formed by splitting the data on environmental values that minimize the intersite SSD.

MRT vs CCA vs RDA

- Unlike CCA/RDA, MRT is *not* a method of direct gradient analysis, in the sense of locating sites along ecological gradients.
  - MRT *does* locate sites (albeit grouped) in an ecological space defined by the environmental variables.

- Unlike CCA/RDA, the MRT solution space is *nonlinear* and includes interactions between the environmental variables.

- MRT is a *divisive partitioning* technique; CCA/RDA model continuous structure.
MRT vs CCA vs RDA

- MRT emphasizes *local structure* and *interactions* between environmental effects; CCA/RDA determine *global structure*.

- MRT does not assume particular relationships between species abundances and environmental characteristics; CCA/RDA assume unimodal and linear response models.

MRT vs CCA vs RDA

- MRT has *outperformed* CCA/RDA in both simulated and real ecological data sets in both explaining and predicting species composition.

- The *advantage* of MRT increases with the strength of interactions and the nonlinearity of relationships between species composition and the environmental variables.