

TOPIC 9

UNCONSTRAINED ORDINATION: CORRESPONDENCE ANALYSIS



Correspondence Analysis (CA)

- see <https://sites.google.com/site/mb3gustame/indirect-gradient-analysis/ca> and https://www.davidzeleny.net/anadat-r/doku.php/en:ca_dca
- also an eigenvector method
- **handles nonlinear species responses better than PCA**
 - **therefore better for community data**
- CA is based on similar regression techniques as PCA, but with χ^2 -standardised data and weights
 - (... i.e. PCA uses an intermediate correlation matrix)
- then subjected to either a Singular Value Decomposition (SVD) or eigenvalue decomposition, and the eigenvalues and eigenvectors reported
- the ordination preserves χ^2 - (D_{16}) rather than Euclidian (D_1) distance between sites
 - χ^2 -distance is not influenced by double 0s
 - no pre-transformation needed
 - **suitable for species counts and presence/absence data**
- CA maximises the correspondence between species scores and sample scores, whereas a PCA maximises the variance explained

Correspondence Analysis (CA)

- CA produces one axis fewer than $\min[n, p]$
- as with PCA, orthogonal axes ranked in **decreasing order of importance**
- the variation represented is the total inertia, **which is the SS of all the values in the χ^2 matrix**
 - (... i.e. not the sum of the eigenvalues along the diagonal as in a PCA)
- individual eigenvalues will always be <1
- **the variation represented along an axis is given by dividing the eigenvalues of the axis by the total inertia**

Correspondence Analysis (CA)

- CA approximates a **unimodal** response model
 - i.e. matches gradients better (fits better to env. data)
 - the species scores give the species maximum and the abundance decreases in every direction from the centroid of the species score
 - (... in PCA species close to the origin (zero) change little and is poorly presented by the ordination, but in CA it may have its optimum there)
- the horseshoe effect is weaker, but still not entirely gone

Correspondence Analysis (CA)

- scaling of ordination plots
 - **scaling 1**—site scaling (rows are the centroids of columns)
 - i.e. sites that plot close together are similar i.t.o. their species relative frequencies
 - any site near a point representing a species will have a relatively large contribution by that species
 - **scaling 2**—species scaling (columns are the centroids of rows)
 - i.e. species points that are close together will have relatively similar frequencies among the objects (sites)
 - any species plotting close-by a point that represents a site is more likely to be found at that object, or have a higher frequency there than at sites appearing further away in the ordination plot

Correspondence Analysis (CA)

```
> spe
# A tibble: 29 x 27
  Cogo  Satr  Phph  Babl  Thth  Teso  Chna  Pato  Lele  Sqce  Baba  Albi  Gogo  Eslu  Pefl  Rham  Legi  Scer  Cyca  Titi
  <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
1     0     3     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
2     0     5     4     3     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0     0
3     0     5     5     5     0     0     0     0     0     0     0     0     0     1     0     0     0     0     0     0
4     0     4     5     5     0     0     0     0     0     1     0     0     1     2     2     0     0     0     0     1
5     0     2     3     2     0     0     0     0     5     2     0     0     2     4     4     0     0     2     0     3
6     0     3     4     5     0     0     0     0     1     2     0     0     1     1     1     0     0     0     0     2
7     0     5     4     5     0     0     0     0     1     1     0     0     0     0     0     0     0     0     0     0
8     0     0     1     3     0     0     0     0     0     5     0     0     0     0     0     0     0     0     0     1
9     0     1     4     4     0     0     0     0     2     2     0     0     1     0     0     0     0     0     0     0
10    1     3     4     1     1     0     0     0     0     1     0     0     0     0     0     0     0     0     0     0
# ... with 19 more rows, and 7 more variables: Abbr <int>, Icme <int>, Gyce <int>, Ruru <int>, Blbj <int>, Alal <int>,
#   Anan <int>
```

```
> (spe.ca <- cca(spe))
Call: cca(X = spe)

              Inertia Rank
Total                1.167
Unconstrained  1.167    26
Inertia is scaled Chi-square

Eigenvalues for unconstrained axes:
  CA1   CA2   CA3   CA4   CA5   CA6   CA7   CA8
0.6010 0.1444 0.1073 0.0834 0.0516 0.0418 0.0339 0.0288
(Shown only 8 of all 26 unconstrained eigenvalues)
```

Correspondence Analysis (CA)

```
> summary(spe.ca)                # default scaling 2

Call:
cca(X = spe)

Partitioning of scaled Chi-square:
      Inertia Proportion
Total      1.167      1
Unconstrained 1.167      1

Eigenvalues, and their contribution to the scaled Chi-square

Importance of components:
      CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8  CA9  CA10  CA11  CA12
Eigenvalue  0.601 0.1444 0.10729 0.08337 0.05158 0.04185 0.03389 0.02883 0.01684 0.010826 0.010142 0.007886
Proportion Explained 0.515 0.1237 0.09195 0.07145 0.04420 0.03586 0.02904 0.02470 0.01443 0.009278 0.008691 0.006758
Cumulative Proportion 0.515 0.6387 0.73069 0.80214 0.84634 0.88220 0.91124 0.93594 0.95038 0.959655 0.968346 0.975104
      CA13  CA14  CA15  CA16  CA17  CA18  CA19  CA20  CA21  CA22  CA23
Eigenvalue  0.006123 0.004867 0.004606 0.003844 0.003067 0.001823 0.001642 0.001295 0.0008775 0.0004217 0.0002149
Proportion Explained 0.005247 0.004171 0.003948 0.003294 0.002629 0.001562 0.001407 0.001110 0.0007520 0.0003614 0.0001841
Cumulative Proportion 0.980351 0.984522 0.988470 0.991764 0.994393 0.995955 0.997362 0.998472 0.9992238 0.9995852 0.9997693
      CA24  CA25  CA26
Eigenvalue  0.0001528 8.949e-05 2.695e-05
Proportion Explained 0.0001309 7.669e-05 2.310e-05
Cumulative Proportion 0.9999002 1.000e+00 1.000e+00

Scaling 2 for species and site scores
* Species are scaled proportional to eigenvalues
* Sites are unscaled: weighted dispersion equal on all dimensions
```

continue...

Correspondence Analysis (CA)

..continue

Species scores

	CA1	CA2	CA3	CA4	CA5	CA6
Cogo	1.50075	-1.410293	0.26049	-0.307203	0.271777	-0.003465
Satr	1.66167	0.444129	0.57571	0.166073	-0.261870	-0.326590
Phph	1.28545	0.285328	-0.04768	0.018126	0.043847	0.200732
Babl	0.98662	0.360900	-0.35265	-0.009021	-0.012231	0.253429
Thth	1.55554	-1.389752	0.80505	-0.468471	0.471301	0.225409
Teso	0.99709	-1.479902	-0.48035	0.079397	-0.105715	-0.332445
Chna	-0.54916	-0.051534	0.01123	-0.096004	-0.382763	0.134807
Pato	-0.18478	-0.437710	-0.57438	0.424267	-0.587150	0.091866
Lele	0.01337	-0.095342	-0.57672	0.212017	0.126668	-0.389103
Sqce	0.01078	0.140577	-0.34811	-0.538268	0.185286	0.167087
Baba	-0.33363	-0.300682	-0.04929	0.170961	-0.157203	0.103068
Albi	-0.38357	-0.255310	-0.20136	0.374057	-0.385866	0.239001
Gogo	-0.32152	-0.034382	-0.07423	-0.031236	0.014417	-0.156351
Eslu	-0.26165	0.187282	0.00617	0.183771	0.295142	-0.262808
Pefl	-0.28913	0.121044	-0.18919	0.367615	0.218087	-0.163675
Rham	-0.60298	-0.057369	0.20341	0.214299	-0.050977	0.211926
Legi	-0.58669	-0.082467	0.21198	0.050175	-0.120456	0.108724
Scer	-0.61815	0.124733	0.13339	0.147190	0.317736	-0.340380
Cyca	-0.57951	-0.110732	0.20173	0.308547	0.006854	0.153224
Titi	-0.37880	0.138023	-0.07825	0.095793	0.256285	-0.029245
Abbr	-0.70235	0.011155	0.40242	0.211582	0.138186	0.132297
Icme	-0.73238	-0.009098	0.55678	0.321852	0.281812	0.172271
Gyce	-0.69300	0.038971	0.37688	-0.183965	-0.051945	-0.011126
Ruru	-0.44181	0.176915	-0.23691	-0.345104	0.129676	-0.043802
Blbj	-0.70928	0.032317	0.40924	0.030224	0.049050	0.114560
Alal	-0.63114	0.053594	0.15204	-0.661381	-0.414796	-0.206611
Anan	-0.63578	-0.041894	0.30093	0.224044	0.030444	0.203160

continue...

Correspondence Analysis (CA)

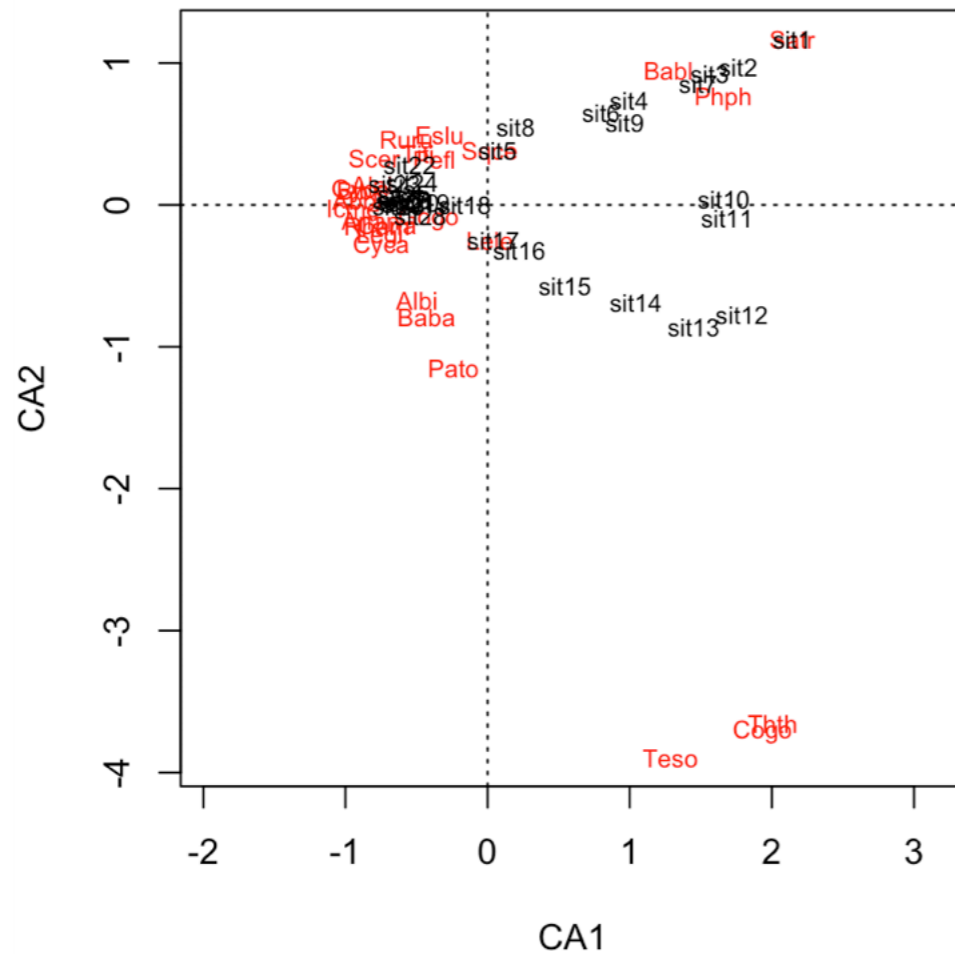
...continue

Site scores (weighted averages of species scores)

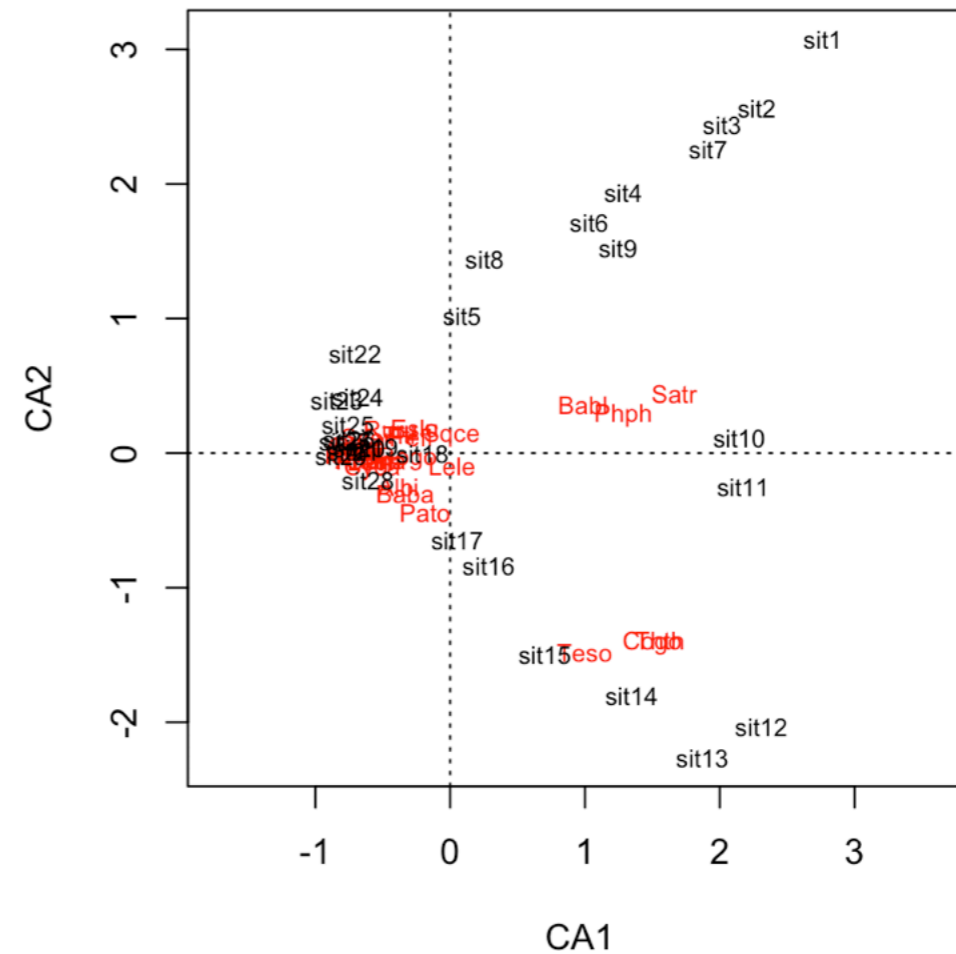
	CA1	CA2	CA3	CA4	CA5	CA6
sit1	2.76488	3.076306	5.3657489	1.99192	-5.07714	-7.80447
sit2	2.27540	2.565531	1.2659130	0.87538	-1.89139	-0.13887
sit3	2.01823	2.441224	0.5144079	0.79436	-1.03741	0.56015
sit4	1.28485	1.935664	-0.2509482	0.76470	0.54752	0.10579
sit5	0.08875	1.015182	-1.4555434	0.47672	2.69249	-2.92498
sit6	1.03188	1.712163	-0.9544059	0.01584	0.91932	0.39856
sit7	1.91427	2.256208	-0.0001407	0.39844	-1.07017	0.32127
sit8	0.25591	1.443008	-2.5777721	-3.41400	2.36613	2.71741
sit9	1.24517	1.526391	-1.9635663	-0.41230	0.69647	1.51859
sit10	2.14501	0.110278	1.6108693	-0.82023	0.53918	1.01153
sit11	2.17418	-0.251649	1.5845397	-0.81483	0.52623	1.05501
sit12	2.30944	-2.034439	1.9181448	-0.60481	0.64435	-0.14844
sit13	1.87141	-2.262503	1.1066796	-0.80840	1.09542	0.11038
sit14	1.34659	-1.805967	-0.6441505	-0.52803	0.76871	-0.67165
sit15	0.70214	-1.501167	-1.9735888	0.98502	-0.93585	-1.27168
sit16	0.28775	-0.836803	-1.2259108	0.73302	-1.57036	0.57315
sit17	0.05299	-0.647950	-0.9234228	0.35770	-0.95401	0.77738
sit18	-0.20584	-0.007252	-1.0154343	0.07041	-1.03450	0.51442
sit19	-0.57879	0.042849	-0.3660551	-0.15019	-0.61357	0.10115
sit20	-0.67320	0.038875	0.1194956	0.17256	-0.14686	-0.12018
sit21	-0.71933	0.014694	0.2204186	0.13598	0.09459	-0.02068
sit22	-0.70438	0.735398	-0.6546250	-6.61523	-2.49441	-1.73215
sit23	-0.83976	0.390120	0.5605295	-4.38864	-2.56916	-0.96702
sit24	-0.68476	0.418842	-0.2860819	-2.80336	-0.37540	-3.93791
sit25	-0.75808	0.210204	0.5894091	-0.70004	-0.01880	-0.10779
sit26	-0.75046	0.100869	0.5531191	-0.12946	0.29164	0.11280
sit27	-0.77878	0.088976	0.7379012	0.05204	0.40940	0.43236
sit28	-0.60815	-0.203235	0.5522726	0.43621	0.15010	0.25618
sit29	-0.80860	-0.019592	0.6686542	0.88136	0.52744	0.16456

Correspondence Analysis (CA)

CA fish abundances - biplot scaling 1



CA fish abundances - biplot scaling 2



- site scaling
- most interested in **sites**
- sites that plot close together are similar i.t.o. their species relative frequencies
- any site near a point representing a species will have a relatively large contribution by that species

- species scaling
- most interested in **species**
- species points that are close together will have relatively similar frequencies among the objects (sites)
- species plotting close-by a point that represents an object (site) are more likely to be found at that object, or have a higher frequency there than at objects appearing further away in the ordination plot

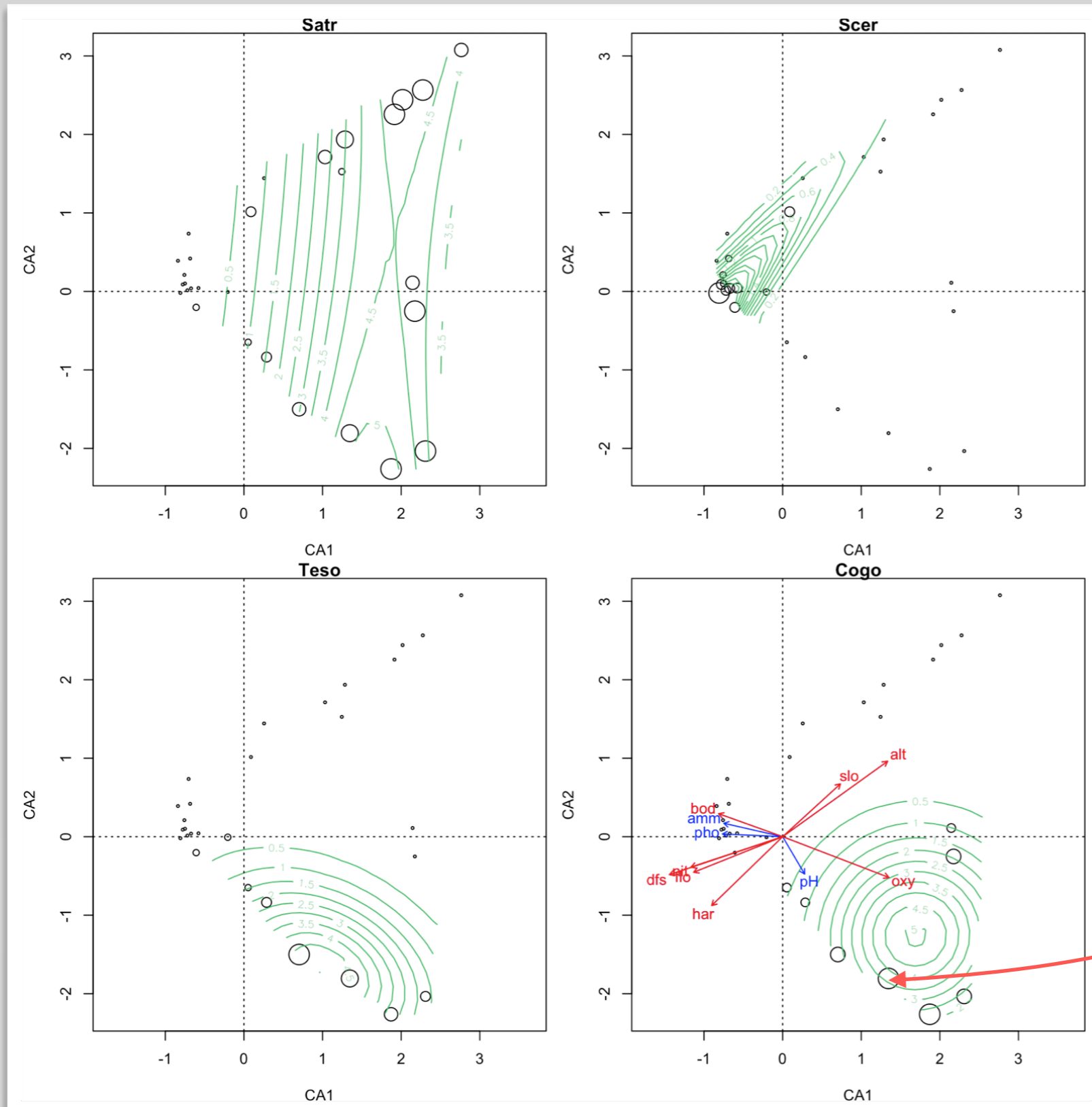
Correspondence Analysis (CA)

Fit and Plot Smooth Surfaces of Variables on Ordination

```
require('viridis')
palette(viridis(8))
par(mar = c(4, 4, 0.9, 0.5) + .1, mfrow = c(2, 2))
with(spe, tmp <- ordisurf(spe.ca ~ Satr, bubble = 3,
                        family = quasipoisson, knots = 2, col = 6,
                        display = "sites", main = "Satr"))
abline(h = 0, v = 0, lty = 3)
with(spe, tmp <- ordisurf(spe.ca ~ Scer, bubble = 3,
                        family = quasipoisson, knots = 2, col = 6,
                        display = "sites", main = "Scer"))
abline(h = 0, v = 0, lty = 3)
with(spe, tmp <- ordisurf(spe.ca ~ Teso, bubble = 3,
                        family = quasipoisson, knots = 2, col = 6,
                        display = "sites", main = "Teso"))
abline(h = 0, v = 0, lty = 3)
with(spe, tmp <- ordisurf(spe.ca ~ Cogo, bubble = 3,
                        family = quasipoisson, knots = 2, col = 6,
                        display = "sites", main = "Cogo"))
abline(h = 0, v = 0, lty = 3)

# A posteriori projection of environmental variables in a CA
# The last plot produced (CA scaling 2) must be active
(spe.ca.env <- envfit(spe.ca, env, scaling = 2)) # Scaling 2 is default
plot(spe.ca.env)
# Plot significant variables with a different colour
plot(spe.ca.env, p.max = 0.05, col = "red")
```

biplot scaling 2



- instead of an arrow of increase, the species score is seen as a centre of abundance
- the species scores give the species maximum and the abundance decreases to every direction from the centroid given by the species score

envfit() not the best option