The fourth corner problem: niche overlap using mixtures

Shirley Pledger

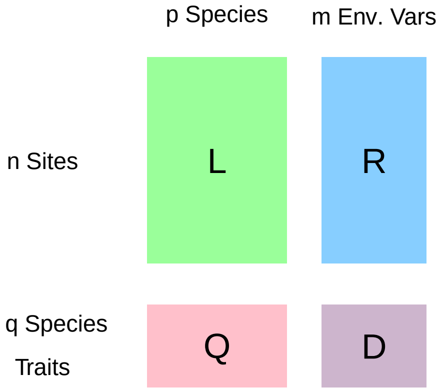
Victoria University of Wellington
New Zealand

IBS Australasian Region Conference
Hobart 30 November 2015

Outline

- 1. The fourth corner problem.
- 2. Data set.
- 3. Interaction plots.
- 4. Niche overlap plots.

1. The Fourth Corner Problem



Have three data matrices:

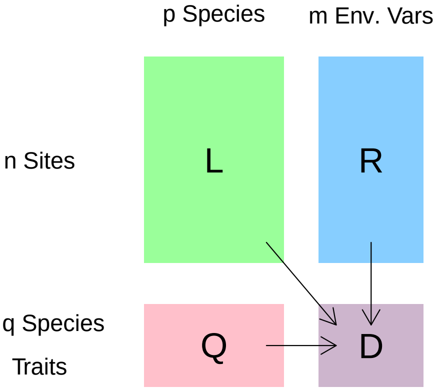
L = abundance or presence/absence

R = environmental data (e.g. altitude, soil type, rainfall)

Q = species traits (e.g. food type, adult size for birds).

Fourth corner is **D** - can we link env. vars. with species traits?

The fourth corner problem has been solved (1)

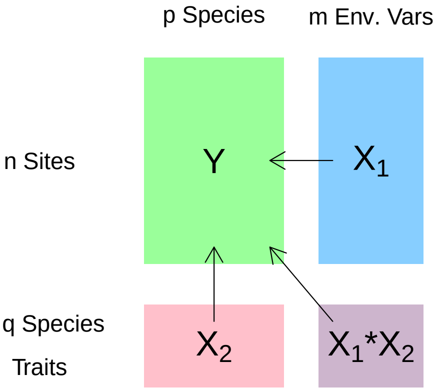


Legendre *et al.* 1997: Matrix algebra, constructed $\mathbf{D} = \mathbf{Q}\ \mathbf{L}^T\ \mathbf{R}$.

L, **R** and **Q** all contribute to estimating **D**.

Revisited 2008, Dray and Legendre, permutation tests.

The fourth corner problem has been solved (2)



Brown 2010, Jamil *et al.* 2013, Brown *et al.* 2014.
$\mathbf{Y}$ is a response matrix, main effects of rows (sites) in $\mathbf{X}_1$, main effects of columns (species) in $\mathbf{X}_2$, interactions in $\mathbf{X}_1 * \mathbf{X}_2$.
Focus on interaction terms, comparisons, testing.

Is there anything left to say?

- **Simplification** of interaction plots.
- **Dimension Reduction** of the response data matrix.
- **Niche Overlap**, a unified measure, plots.

We need a data set to illustrate the models.

2. Data Set: AVIURBA Bird Count Data

Tatibouet, 1981. Thesis, University of Lyons. R package ade4.
Details: n = 51 sites (relevés), p = 40 spp. of birds.

Eleven environmental variables:
One is percentage vegetation cover.
Ten are presence/absence: farms, small.buildings, high.buildings, industry, fields, grassland, scrubby, deciduous, conifer, noisy.

Four species traits:
Feeding habit (3 levels, insectivore, granivore, omnivore); Feeding stratum (3 levels, ground, aerial, scrub/foliage); Breeding stratum (4 levels, ground, building, scrub, foliage); Migratory stratum (2 levels, resident, migrant).

AVIURBA Count Matrix

R1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	0	0	0	1	0	0	4	0	0	1	1	0	1	0	0	0	0	1
R2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2	1	0	0	1	0	1	0	1	0	0	4	4	1	1	1	0	0	0	0	0	0	0
R3	0	2	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	2	0	0	0	1	0	0	0	1	1	0	1	0	1	0	1	0	0	0	0	0	0	1
R4	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	2	0	0	0	1	0	1	0	0	2	0	1	0	1	1	1	0	0	0	0	1	0	0
R5	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0
R6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	1	1	2	1	2	0	0	0	0
R7	0	0	0	4	0	0	0	0	1	1	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	4	1	0	1	1	0	0	1	0	0	0	0
R8	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	2	1	0	1	1	1	1	1	1	0	0	1	0	1	0	0	0	1	1	1	0	0	1
R9	0	0	0	1	0	1	0	1	0	4	4	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	0	0	3	2	1	1	0	0	0	2	0	0	0	0
R10	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	2	1	1	1	0	1	0	1	1	2	1	1	1	0	0	1	0	0	1
R11	0	1	0	0	0	1	1	3	0	1	0	0	0	0	0	0	0	1	2	0	0	1	0	0	0	1	0	1	3	1	0	1	1	1	1	2	1	0	0	0
R12	1	0	0	0	0	0	1	2	1	0	1	1	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	4	0	0	1	1	0	1	0	0	0	0	1
R13	0	0	0	0	0	0	0	0	1	1	0	1	0	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
R14	0	0	0	0	0	0	0	3	0	0	3	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	2	1	0	1	0	0	0	0	0
R15	0	0	0	2	0	1	1	2	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	1	0	1
R16	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	3	0	1	2	0	0	1	1	0	2	0	1
R17	0	0	0	3	0	0	0	3	0	2	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	3	2	0	1	0	0	1	0	0	1	0	1
R18	0	0	0	0	0	1	0	3	0	0	3	0	0	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	3	2	0	2	0	0	1	1	0	0	0	1
R19	0	1	0	0	0	1	0	3	1	1	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	3	0	0	1	0	0	1	0	0	0	0	0
R20	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	2	1	0	1	0	0	1	0	0	0	1	0
R21	0	0	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	1	3	0	0	3	0	0	0	0	0
R22	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	1	0	0
R23	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	1	0
R24	0	0	0	1	0	1	0	3	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0
R25	0	0	0	3	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	1	0	0	0	0	1
R26	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	1	0	0	3	0	1	1	0	0	1	0	0	0	0	0
R27	0	0	1	3	0	0	0	2	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	3	0	0	2	0	0	1	2	0	0	0	0
R28	0	0	0	3	0	1	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	1	0	0	0	0	0
R29	0	0	0	3	0	1	0	3	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	1	0	0	0	0	0
R30	0	1	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0
R31	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0
R32	0	0	0	0	0	0	0	3	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0
R33	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	1	0	1	0	0	1	2	0	0	0	2
R34	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	1	0	0	2	1	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	1	0	0
R35	0	0	0	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0
R36	0	0	0	1	0	1	0	4	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	4	0	0	1	0	0	1	0	0	0	0	0
R37	0	0	0	1	0	1	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0
R38	0	0	0	0	0	0	0	3	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0
R39	0	0	0	2	0	0	0	3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	1	0	0	1	3	0	1	0	0
R40	0	0	0	1	0	0	0	3	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	1	0	1	0	0
R41	0	0	0	0	0	0	0	3	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	3	3	0	2	0	0	1	0	0	0	0	0
R42	0	1	0	0	0	0	0	3	1	3	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	3	1	0	1	0	0	1	1	0	0	0	0
R43	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	0
R44	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
R45	0	1	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	3	0	0	1	1	0	1	0	0	0	0	2
R46	1	0	0	0	0	0	0	3	2	3	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0
R47	0	0	0	0	0	0	0	3	1	3	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	4	4	0	1	0	0	1	0	0	0	0	0
R48	0	0	1	0	0	0	0	3	3	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	1	0	0	0	1	0	0	0	3
R49	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
R50	0	0	1	0	2	0	1	0	3	2	2	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	2	0	0	2	0	1	0	0	0	0	0	0	0	0
R51	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	3	0	0	1	1	0	0	2	0	0	0	1

y	frequency
0	1548
1	336
2	63
3	80
4	13
	2040

Mean $= 0.37$

3.Interaction Plots

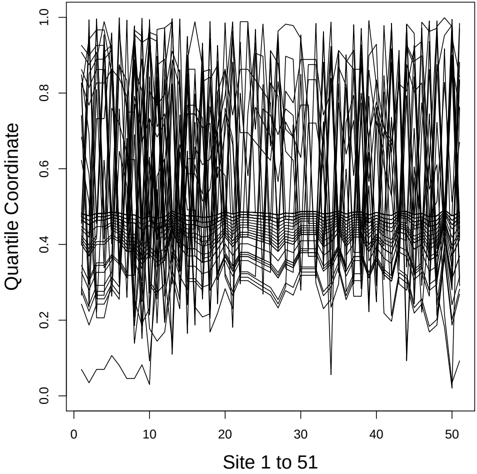
Generalized linear models (GLMs):
$E(Y_{ij}) = \mu_{ij}$. Link function $g(.)$, $\quad g(\mu_{ij}) = \eta_{ij} =$ linear predictor.

Saturated model:
$\eta_{ij} = \nu + \alpha_i + \beta_j + \gamma_{ij} =$ Intercept + Row + Col. + Row×Col.
Constraints: e.g. $\sum \alpha_i = 0$, $\sum \beta_j = 0$, $\forall i$, $\sum_j \gamma_{ij} = 0$, $\forall j$, $\sum_i \gamma_{ij} = 0$.

No-association model: (Main effects only)
$\eta_{ij} = \nu + \alpha_i + \beta_j, \quad \sum \alpha_i = 0, \sum \beta_j = 0.$

Note: Inclusion of α_i and β_j is like standardising over rows (sites) and columns (species). It prevents the analysis being dominated by numerous species or by rich sites, and allows γ to focus on associations and turnover patterns.

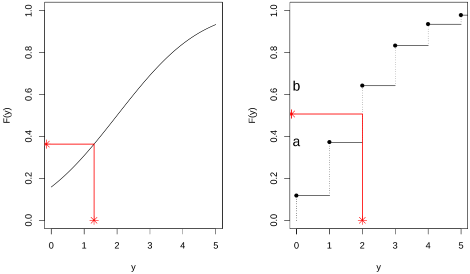
Interaction Plot for Saturated Model



40 species,
One trace per species.

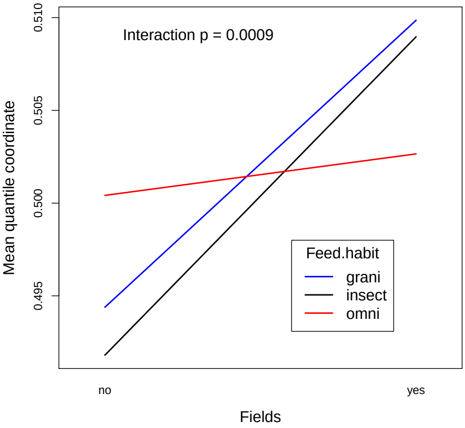
Y axis is quantile coordinate,
a probability integral transform;
represents deviation from no-assoc. model.

Quantile Coordinates (Prob. Integral Transforms)



Pattern: in residuals after fitting no-association model.
Which residuals? Problem with log scale: log(0) if $y_{ij} = 0$.
Problem with raw data scale: large counts create asymmetry of $y - \widehat{y}$ or $y/\widehat{y}$.
Use cdf scale. If no-association model fits well, randomised quantile residuals $\sim$ Unif[0,1] (diagnostic Q-Q Unif plots).
Pattern is shown by quantile coordinate plots.

Interaction of Two Covariates



Model $g(E(Y)) = \text{Intercept} + \text{Site} + \text{Species} + \text{Fields}\times\text{Feed.habit}$

Cluster sites and/or species by finite mixtures

Finite mixtures: still likelihood-based methodology.

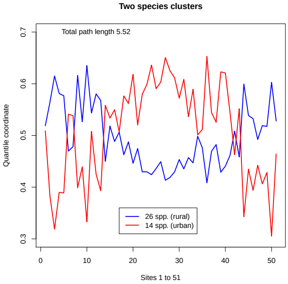
Row clustering: Cluster sites, keep all species separate.
$g(E(Y_{ij})) = \nu + \alpha_i + \beta_j + \gamma_{rj} \quad r = 1, \ldots, R.$

Column clustering: Cluster species, keep all sites separate
$g(E(Y_{ij})) = \nu + \alpha_i + \beta_j + \gamma_{ic} \quad c = 1, \ldots, C.$

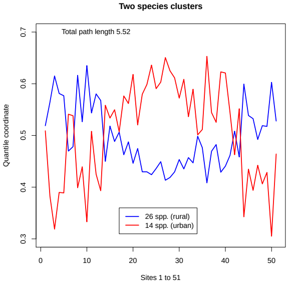
Biclustering: Simultaneously cluster sites and species.
$g(E(Y_{ij})) = \nu + \alpha_i + \beta_j + \gamma_{rc}.$

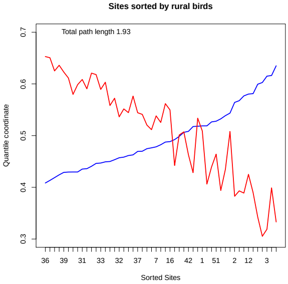
Unsupervised learning. EM algorithm. Posterior probabilities (fuzzily) allocate sites to site clusters, species to species clusters. Next, look at plots from column (species) clustering, all sites separate.

Interaction: All Sites × Two Species Clusters γ_{ic}

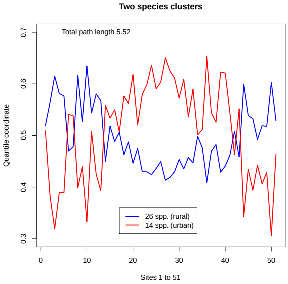


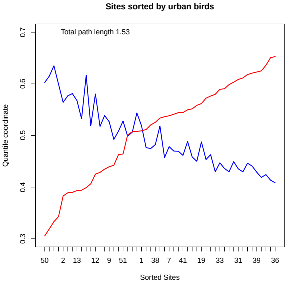
Interaction: All Sites $\times$ Two Species Clusters γ_{ic}

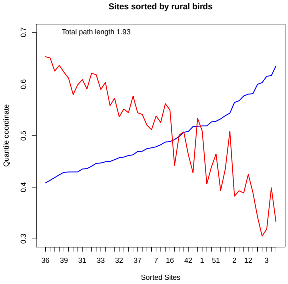




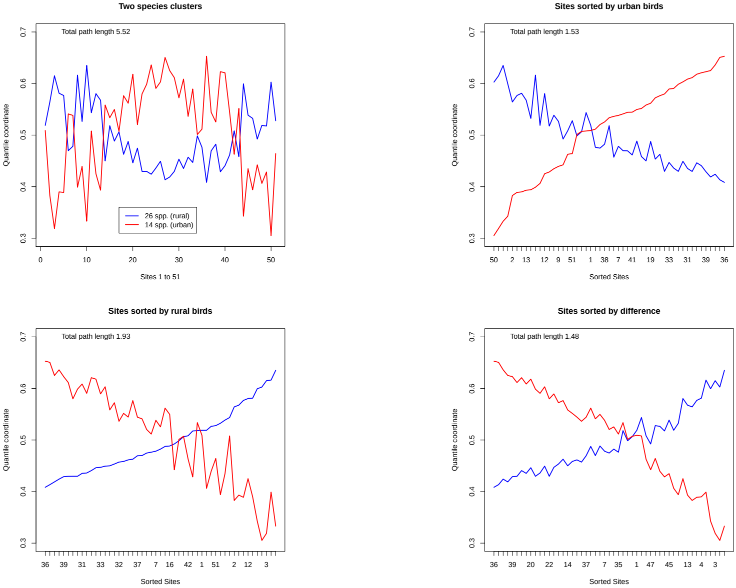
Interaction: All Sites $\times$ Two Species Clusters γ_{ic}



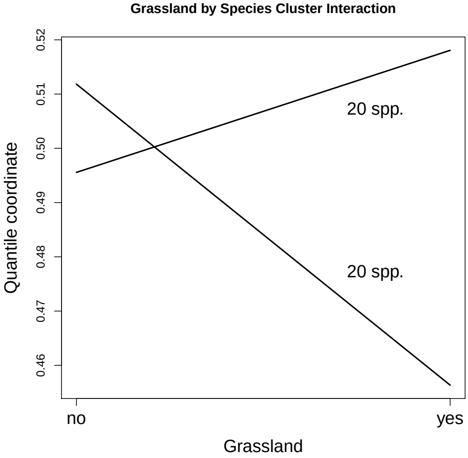




Interaction: All Sites × Two Species Clusters γ_{ic}

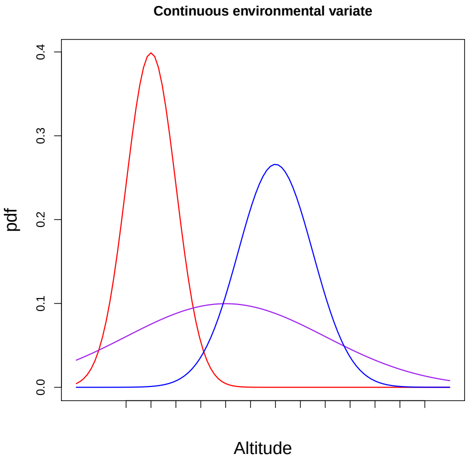


Hybrid Model: Site Variable by Species Cluster



$$g(E(Y)) = \text{Intercept} + \text{Site} + \text{Species} + \text{Grass}\times\text{Sp.Clust.}$$

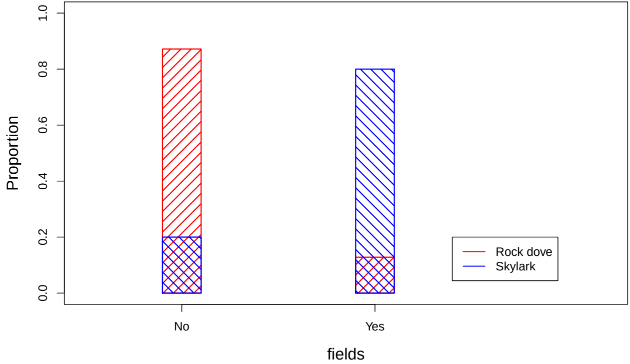
4. Niche Overlap Plots



Continuous variates, overlap of pdf. Density estimation.

$0 \leq$ niche overlap ≤ 1. Niche separation $=$ 1 - niche overlap.

Fields Niche Overlap for Two Species



Categorical environmental variable.
Niche overlap = double-shaded area = 0.32.
Lower than random, so significant niche separation ($p = 0.000$).

Other Types of Environmental Variates or Traits

Geange *et al.* 2011. Unified models for different types of variates.

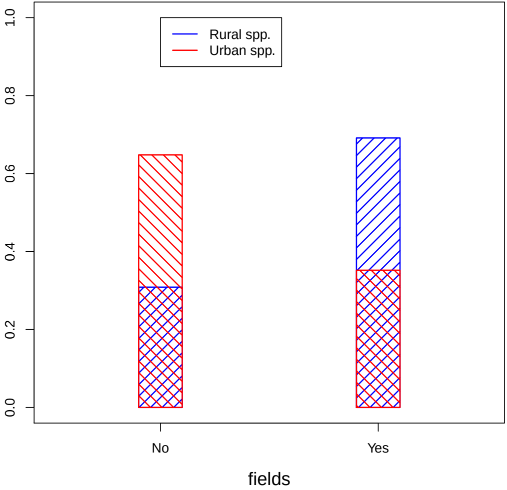
Overlap of two probability distributions for continuous, positive continuous, percentage, categorical (binomial or multinomial) or count data.

Example: Over K categories

Niche overlap of i and j is $\sum_{k=1}^{K} min(p_{ik}, p_{jk})$.

Niche separation = 1 - niche overlap = $\sum_{k=1}^{K} \mid p_{ik} - p_{jk} \mid$.

Niche Overlap of Two Species Archtypes from Column Clustering



Niche overlap = 0.661, p = 0.000, Significant niche separation.

Using More Information

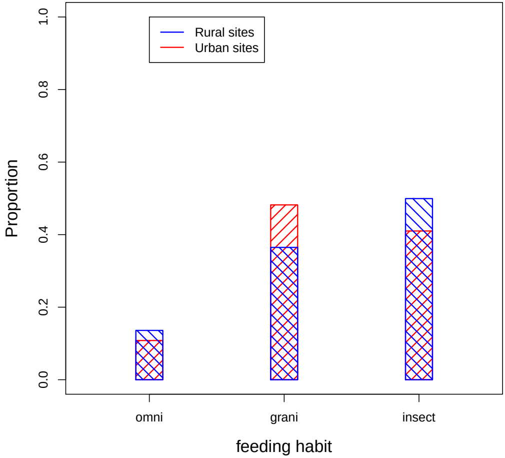
We have fourth corner information, including Species Traits.

These may also be used for niche overlap, by transposing the ideas.

Use two site clusters from a row-clustered model, "rural" and "urban", say.

Which niches do the site clusters occupy in relation to (say) feeding habit (3 levels, insectivore, granivore, omnivore)?

Do site types have niches in the birds they attract?



Niche overlap = 0.883. No significant separation (p = 0.550).

Work in Progress

With Daniel Fernández, exploring niche overlap methods for data clustered by finite mixtures.

The abundance measure in matrix Y may be binary, count, continuous or ordinal categorical data.

Thank you for coming!

References

Brown, A. (2010). The fourth corner problem: a model-based approach to grassland Hemipteran assemblages. Honours Thesis, University of N.S.W.

Brown, A., Warton, D., Andrew, N., Binns, M., Cassis, G. and Gibb, H. (2014). The fourth corner solution - using predictive models to understand how species traits interact with the environment. *Methods in Ecology and Evolution*, 5: 344–352.

Dray, S. and Legendre, P. (2008). Testing the species traits-environment relationships: the fourth cornere problem revisited. *Ecology*, 89, 3400–3412.

Geange, S., Pledger, S., Burns, K. and Shima, J. (2011). A unified analysis of niche overlap incorporating data of different types. *Methods in Ecology and Evolution*, 2, 175-184.

References (continued)

Jamil, T., Ozinga, W., Kleyer, M. and ter Braak, C. (2013). Selecting traits that explain species-environment relationships: a generalized linear mixed model approach. *Journal of vegetation science*, 24, 988–1000.

Legendre, P., Galzin, R. and Hamelin-Vivien, M. (1997). Relating behaviour to habitat: solutions to the fourth-corner problem. *Ecology*, 78, 547–562.

Pledger, S. and Arnold, R. (2014). Multivariate methods using mixtures: Correspondence analysis, scaling and pattern detection. *Computational Statistics and Data Analysis*, 71: 241-261.

Tatibouet, F. (1981). Approche écologique d'un établissment humain (environement et structure). Example de la communauté urbained de Lyon, University of Lyon.