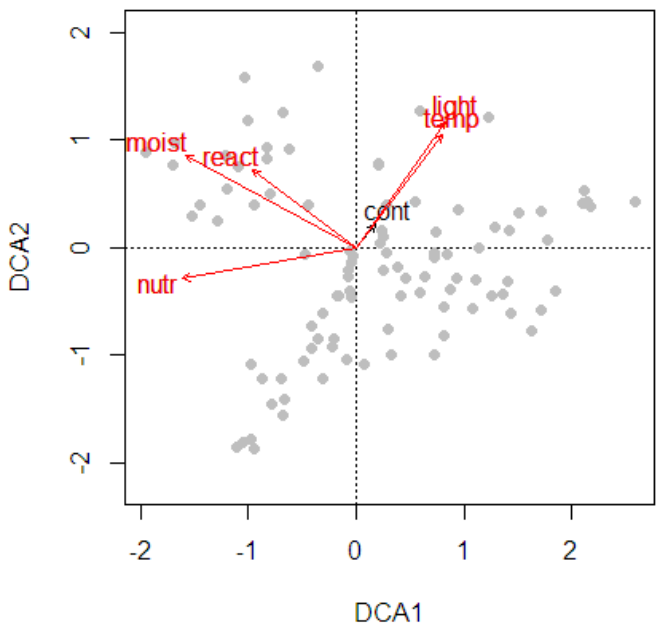
Bias in community-weighted mean analysis of plant functional traits and species indicator values

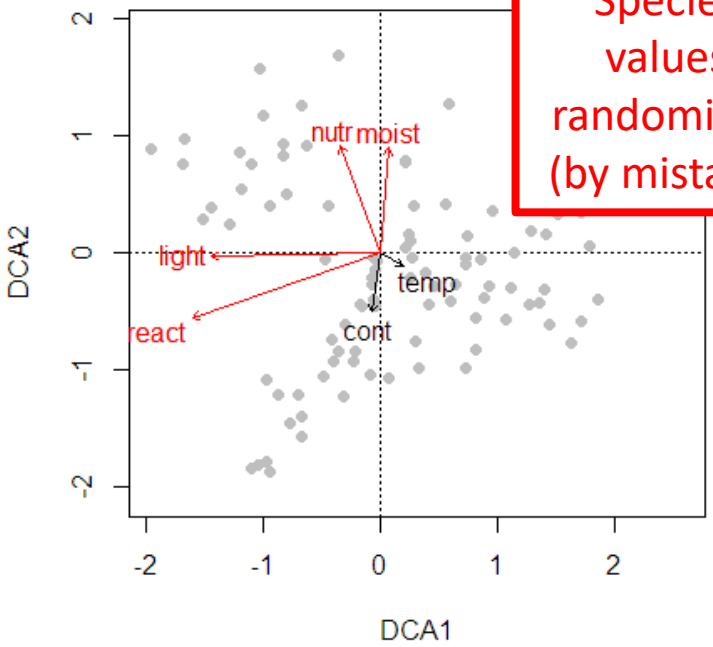
David Zelený

Institute of Ecology and Evolutionary Biology
National Taiwan University
Taipei, Taiwan

Mean Ellenberg indicator values as supplementary variables in DCA



	DCA1	DCA2	r2	Pr(>r)	
light	0.57495	0.81819	0.5187	**0.001**	***
temp	0.60487	0.79633	0.4351	**0.001**	***
cont	0.64357	0.76539	0.0194	0.429	
moist	-0.88038	0.47428	0.8149	**0.001**	***
react	-0.79798	0.60269	0.3637	**0.001**	***
nutr	-0.98460	-0.17482	0.6727	**0.001**	***



Species values randomized (by mistake)

	DCA1	DCA2	r2	Pr(>r)	
light	-0.99979	-0.02043	0.2837	**0.001**	***
temp	0.85009	-0.52663	0.0069	0.769	
cont	-0.15721	-0.98756	0.0356	0.246	
moist	0.07190	0.99741	0.1100	**0.008**	**
react	-0.94473	-0.32785	0.3938	**0.001**	***
nutr	-0.35387	0.93530	0.1308	**0.004**	**

~ 67% probability of result significant at P < 0.05

Journal of Vegetation Science **23** (2012) 419–431

Too good to be true: pitfalls of using mean Ellenberg indicator values in vegetation analyses

David Zelený & André P. Schaffers

Keywords
Bio-indication; Circularity of reasoning; Compositional similarity; Environmental calibration; Null models; Ordination analysis; Randomization; Species richness; Weighted averaging

Abbreviations
ANOVA = analysis of variance; CA = correspondence analysis; CCA = canonical correspondence analysis; DCA = detrended correspondence analysis; EIVs = Ellenberg indicator values; TWINSPAN = two-way indicator species analysis

Received 2 September 2010
Accepted 14 October 2011
Co-ordinating Editor: Michael Palmer

Zelený, D. (corresponding author,

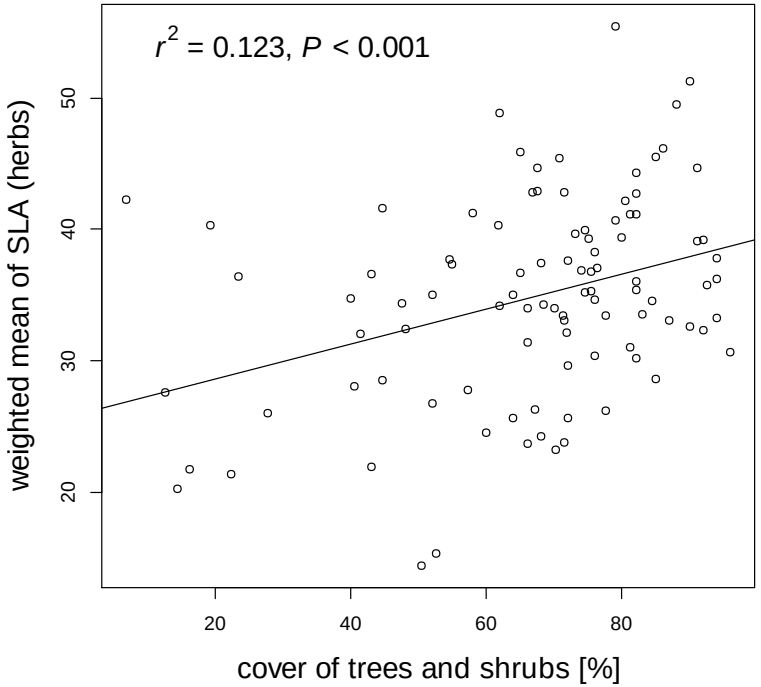
Abstract

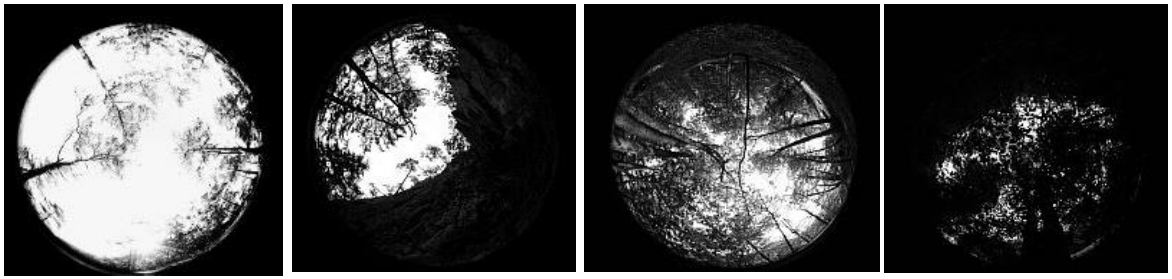
Question: Mean Ellenberg indicator values (EIVs) inherit information about compositional similarity, because during their calculation species abundances (or presence–absences) are used as weights. Can this *similarity issue* actually be demonstrated, does it bias results of vegetation analyses correlating mean EIVs with other aspects of species composition and how often are biased studies published?

Methods: In order to separate information on compositional similarity possibly present in mean EIVs, a new variable was introduced, calculated as a weighted average of randomized species EIVs. The performance of these *mean randomized EIVs* was compared with that of the *mean real EIVs* on the one hand and *random values* (randomized mean EIVs) on the other. To demonstrate the *similarity issue*, differences between samples were correlated with dissimilarity matrices based on various indices. Next, the three mean EIV variables were tested in canonical correspondence analysis (CCA), analysis of variance (ANOVA) be species richness. Subsequently, a

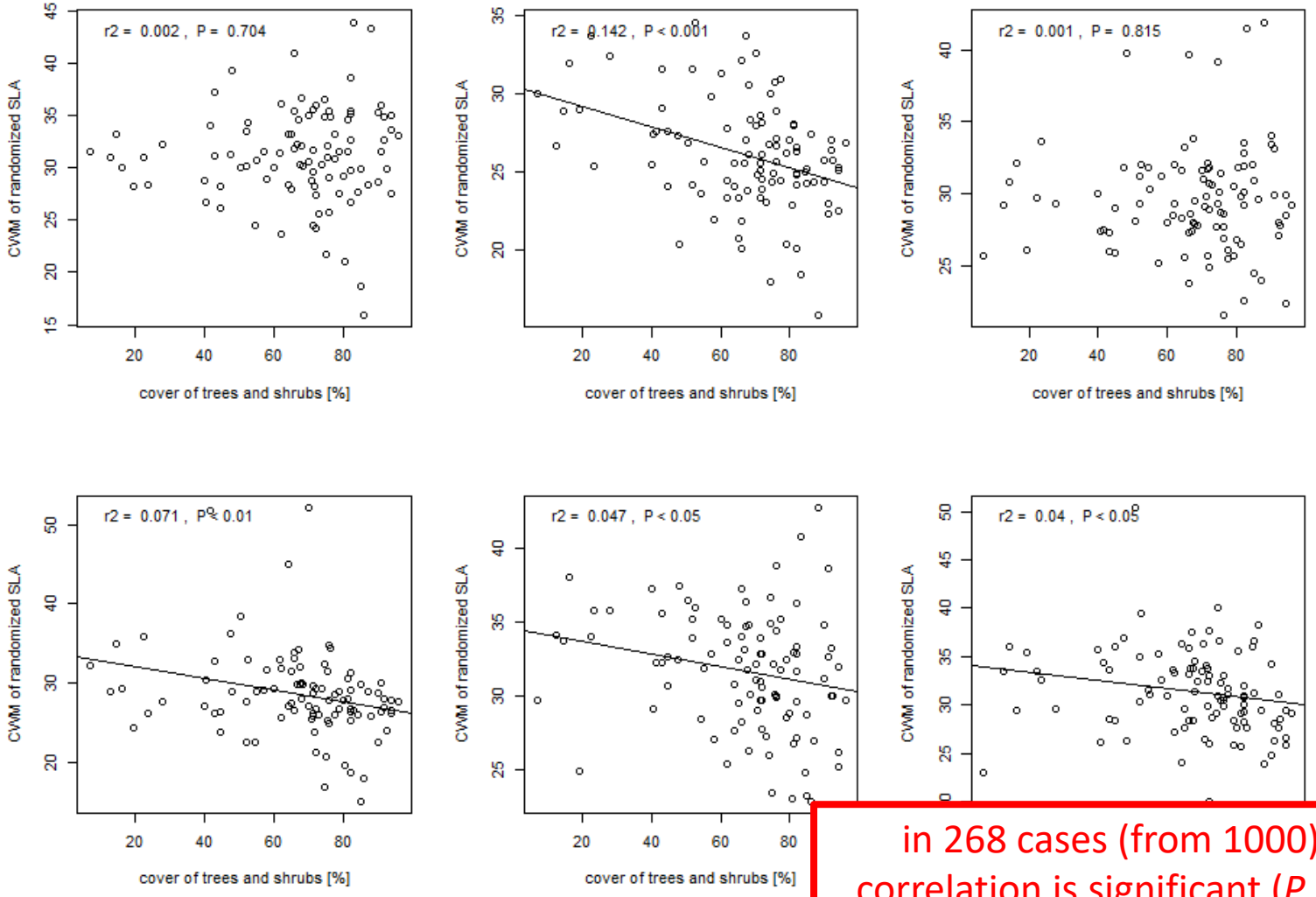
Cited by: WoS: 72, Scopus: 76, Google Scholar: 98

CWM of specific leaf area (SLA) related to tree cover (Vltava valley data)



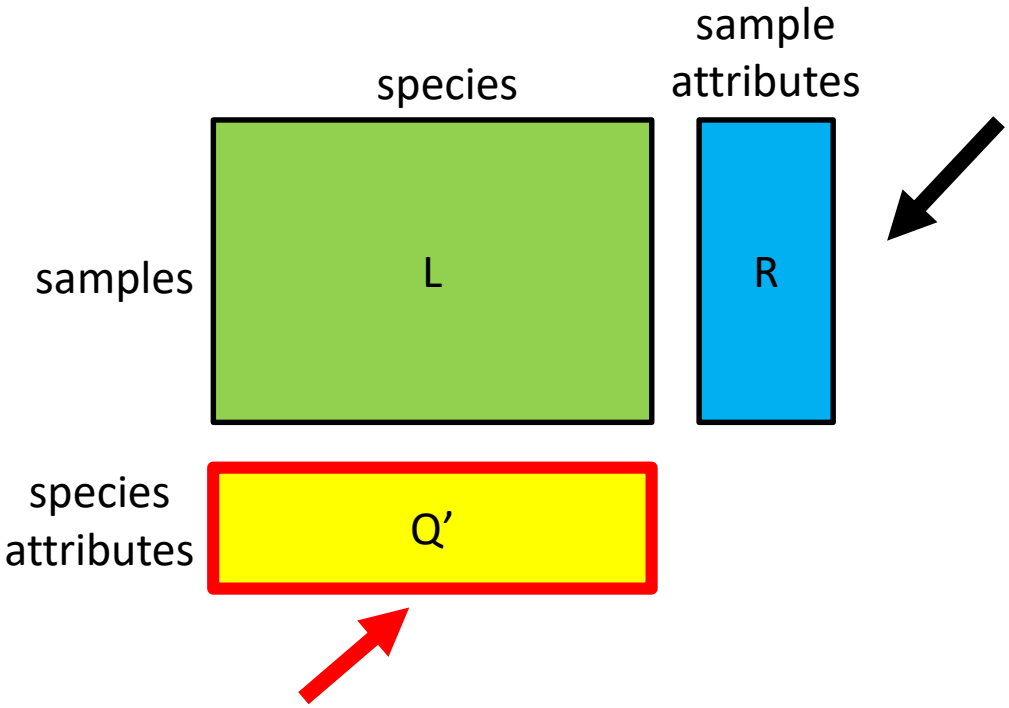


Relationship of CWM on randomized species attributes with environmental variable (tree cover)



in 268 cases (from 1000) the correlation is significant ($P < 0.05$)

Analysis of species and sample attributes



sample attributes – characteristics of individual samples (plots, localities)

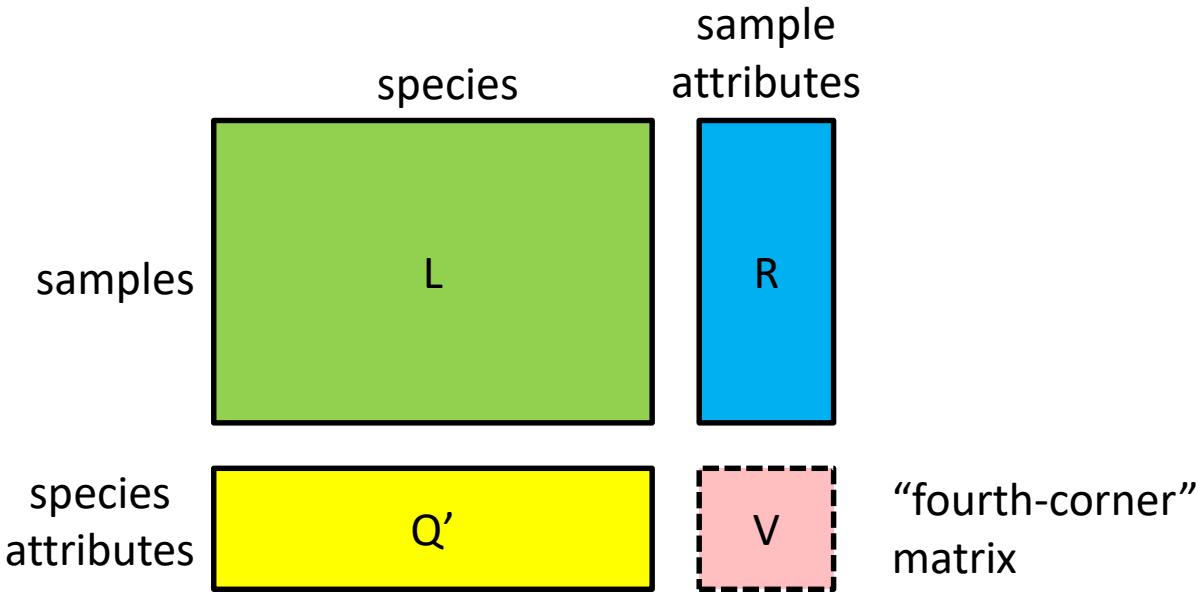
- environmental variables
- experimental treatment
- group assignments (classification)
- diversity
- ordination scores

species attributes – characteristics of individual species

- traits
- species indicator values
- species phylogenetic relationship
- species specialization values etc.

Intrinsic vs *extrinsic* sample or species attributes

Analysis of species and sample attributes



Three solutions:

1) combining (**L** & **Q**) and relate to **R** - *community-weighted mean* approach
2) combining (**L** & **R**) and relate to **Q** - *species centroid* approach
3) simultaneous **R** & **L** & **Q** analysis - *fourth-corner* approach, RLQ, GLM, GLMM

Problem

- If the relationship between CWM and sample attributes (e.g. environmental variables) is tested by standard tests, the Type I error is inflated and effect size (like r or r^2) has higher variance or is biased

- This problem is contingent upon beta diversity of the species composition dataset, and the strength of the relationship between sample attributes and species composition

100 YEARS Journal of Ecology

British Ecological Society

Journal of Ecology doi: 10.1111/1365-2745.12065

Trait-mediated effects of environmental filtering on tree community dynamics

Jesse R. Lasky[1], I-Fang Sun[2]*, Sheng-Hsin Su[3], Zueng-Sang Chen[4] and Timothy H. Keitt[1]

Table 2. Estimated slopes of general linear models, which relate interspecific mean trait values in quadrats (weighted by stem frequency) to quadrat environment

	Convexity (m)	*P*	Slope (10°)	*P*	Ava. N (10 mg kg^{-1})	*P*	Ava. P (mg kg^{-1})	*P*
Leaf area [log (cm^2)]	−0.040	< 0.0001	0.000	0.1999	−0.020	< 0.0001	0.016	0.0382
Specific leaf area [log (cm^2 g^{-1})]	−0.010	< 0.0001	−0.017	< 0.0001	−0.001	0.3482	0.007	0.0336
Leaf succulence [log (mg cm^{-2})]	0.000	0.6887	0.007	0.0024	−0.002	0.0169	0.005	0.0141
Max. height [log (m)]	0.010	0.0025	0.012	0.0354	0.001	0.5961	−0.024	< 0.0001
Wood density (g cm^{-3})	0.005	< 0.0001	0.003	0.0019	0.001	0.0003	−0.002	0.0501
PC1	−0.034	< 0.0001	−0.061	< 0.0001	−0.004	0.4567	0.028	0.0357
PC2	0.026	< 0.0001	0.009	0.1953	0.016	< 0.0001	0.009	0.1207

Slopes are given in units of trait/environment. *P*-values show results for testing the null hypothesis that slope = 0.

Ecology, 89(12), 2008, pp. 3400–3412
© 2008 by the Ecological Society of America

TESTING THE SPECIES TRAITS–ENVIRONMENT RELATIONSHIPS: THE FOURTH-CORNER PROBLEM REVISITED

STÉPHANE DRAY[1,3] AND PIERRE LEGENDRE[2]

[1]*Univ*

[2]*Département de*

and
Fro
data
of s
tabl
year
betv
prac
resti
anal
ofte
this
appr

Ecology, 93(7), 2012, pp. 1525–1526
© 2012 by the Ecological Society of America

Improved testing of species traits–environment relationships in the fourth-corner problem

CAJO J. F. TER BRAAK,[1,4] ANOUK CORMONT,[2] AND STÉPHANE DRAY[3]

[3]*Universi*

Ecology, 95(1), 2014, pp. 14–21
© 2014 by the Ecological Society of America

Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation

STÉPHANE DRAY,[1,9] PHILIPPE CHOLER,[2,3] SYLVAIN DOLÉDEC,[4] PEDRO R. PERES-NETO,[5] WILFRIED THUILLER,[2] SANDRINE PAVOINE,[6,7] AND CAJO J. F. TER BRAAK[8]

[1]*Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France*
[2]*Laboratoire d'Ecologie Alpine, UMR CNRS-UJF 5553, Université J. Fourier, BP 53, F-38041 Grenoble, France*
[3]*Station Alpine J. Fourier, UMS CNRS-UJF 3370, Université J. Fourier, BP 53, F-38041 Grenoble, France*
[4]*Université de Lyon, F-69000 Lyon, Université Lyon 1, CNRS, UMR5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, F-69622 Villeurbanne, France*
[5]*Canada Research Chair in Spatial Modelling and Biodiversity, Département des Sciences Biologiques, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, Montréal, Quebec H3C 3P8 Canada*

Ecography 39: 001–011, 2016
doi: 10.1111/ecog.02302
© 2016 The Authors. Ecography © 2016 Nordic Society Oikos
Subject Editor: Miguel Nakamura. Editor-in-Chief: Miguel Araújo. Accepted 5 July 2016

Linking trait variation to the environment: critical issues with community-weighted mean correlation resolved by the fourth-corner approach

Pedro R. Peres-Neto, Stéphane Dray and Cajo J. F. ter Braak

P. R. Peres-Neto (peres-neto.pedro@uqam.ca), Ca
du Québec à Montréal, Succursale Centre-Ville
Univ. Lyon 1, CNRS, UMR 5558, Villeurbann
the Netherlands.

Establishing trait–environment relatio
community weighted means correlatio
the species niche centroid correlation
Using mathematical derivations and si
and unacceptable significance rates wh
themselves are not linked. We show ho
replace all by the Chessel fourth-corne
able value. We propose an appropriat
statistical power in detecting trait–envi
partitioned among and within commu
a contrast between methods and an ap

PeerJ

A critical issue in model-based inference for studying trait-based community assembly and a solution

Cajo J.F. ter Braak[1], Pedro Peres-Neto[2] and Stéphane Dray[3]

[1] Biometris, Wageningen University & Research, Wageningen, The Netherlands
[2] Department of Biology, Concordia University, Montreal, Canada
[3] Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard (Lyon I), Villeurbanne, France

ABSTRACT

Statistical testing of trait-environment association from data is a challenge as there is no common unit of observation: the trait is observed on species, the environment on sites and the mediating abundance on species-site combinations. A number of correlation-based methods, such as the community weighted trait means method (CWM), the fourth-corner correlation method and the multivariate method RLQ, have been proposed to estimate such trait-environment associations. In these methods, valid statistical testing proceeds by performing two separate resampling tests, one site-based and the other species-based and by assessing significance by the largest of the two

Journal of Biogeography (J. Biogeogr.) (2017)

Structural bias in aggregated species-level variables driven by repeated species co-occurrences: a pervasive problem in community and assemblage data

Bradford A. Hawkins[1]*, Boris Leroy[2], Miguel Á. Rodríguez[3], Alexander Singer[4,5,6], Bruno Vilela[3,7], Fabricio Villalobos[7,8], Xiangping Wang[9] and David Zelený[10]

[1]*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA,* [2]*UMR 7208 BOREA, Muséum National d'Histoire Naturelle, Sorbonne Universités, Université Pierre et Marie Curie, Université de Caen, CNRS, IRD, Université des Antilles, Paris, France,* [3]*Forest Ecology & Restoration Group, Department of Life Sciences, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain,* [4]*German Centre for Integrative Biodiversity Research (iDiv)*

ABSTRACT

Aim Species attributes are often used to explain diversity patterns across assemblages/communities. However, repeated species co-occurrences can generate spatial pattern and strong statistical relationships between aggregated attributes and richness in the absence of biological information. Our aim is to increase awareness of this problem.

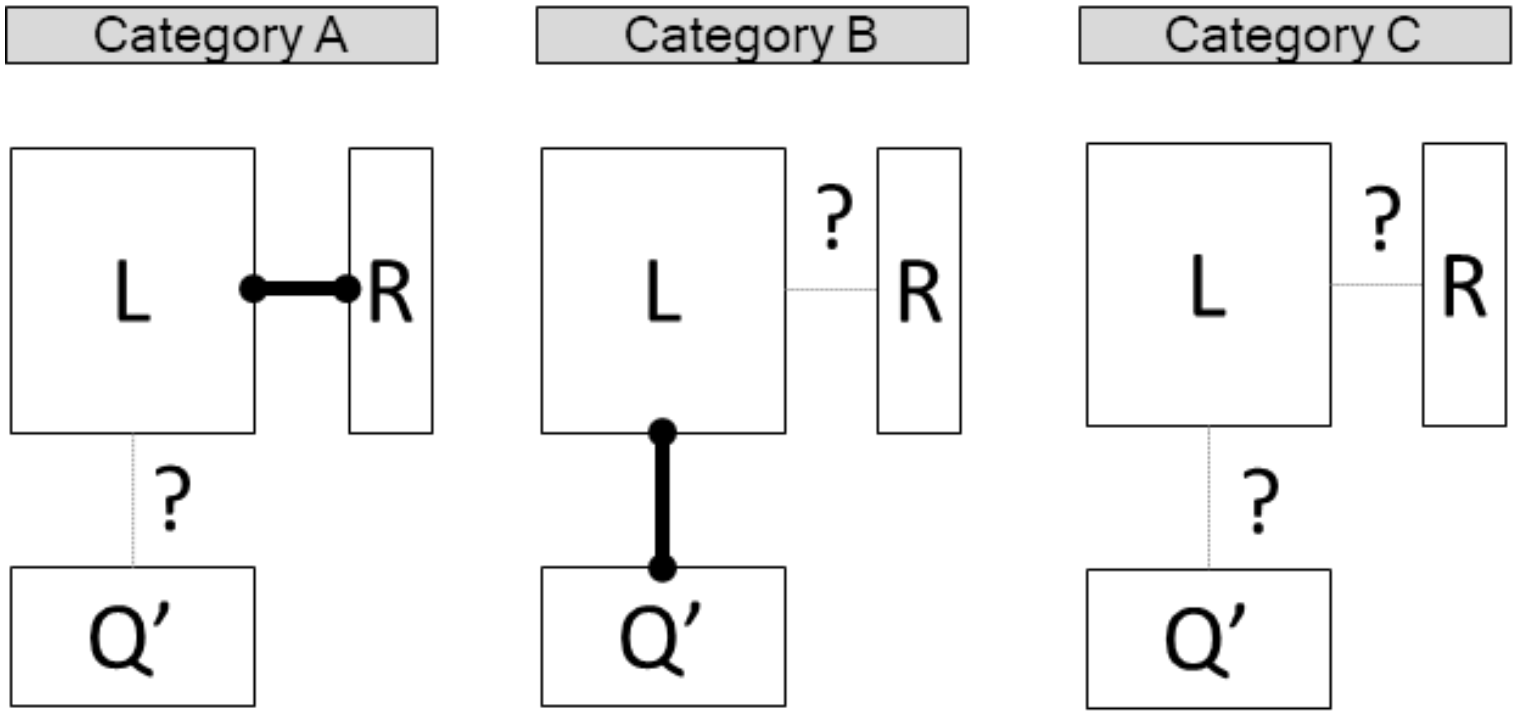
Location North America.

Methods We generated empirical species richness patterns using two data structures: (1) birds gridded from range maps and (2) tree communities from

Progress so far

- It seems that there is a general analytical tool for analysis of CWM with sample attributes – weighted regression (equal to *fourth-corner statistic*) with sequential permutation test (Peres-Neto et al. 2017); it does not return biased Type I error rate and has high power
- It remains to be evaluated how general this approach really is (e.g. if it can work also without weighting the sample attributes by sample totals)
- **But – what to do with all the studies which are already published?**

Proposal: to introduce three categories of hypotheses and identify suitable test for each



Explanation: ⊷ = an assumption of the relationship, **?** = what is being tested

Proposal: introduce three categories of hypotheses and identify suitable test

Category A

- Relationship of CWM to *extrinsic* sample attributes – e.g. experimental treatment (we assume that treatment is effective, and evaluate how traits of species indicators react to it)
- Relationship of CWM to *intrinsic* sample attributes – e.g. sample scores in unconstrained ordination, assignment of samples into groups by cluster analysis, but also diversity

Standard test has inflated Type I error rate (too good to be true)

Proposal: introduce three categories of hypotheses and identify suitable test

Category B

- Studies assuming the link of species attributes to species composition and testing the effect of sample attributes
- e.g. trait studies assuming that measured traits are functional, i.e. they influence species abundance and are linked to species composition, and testing whether environmental variables act as an environmental filter on species abundances.

Standard test has correct Type I error rate

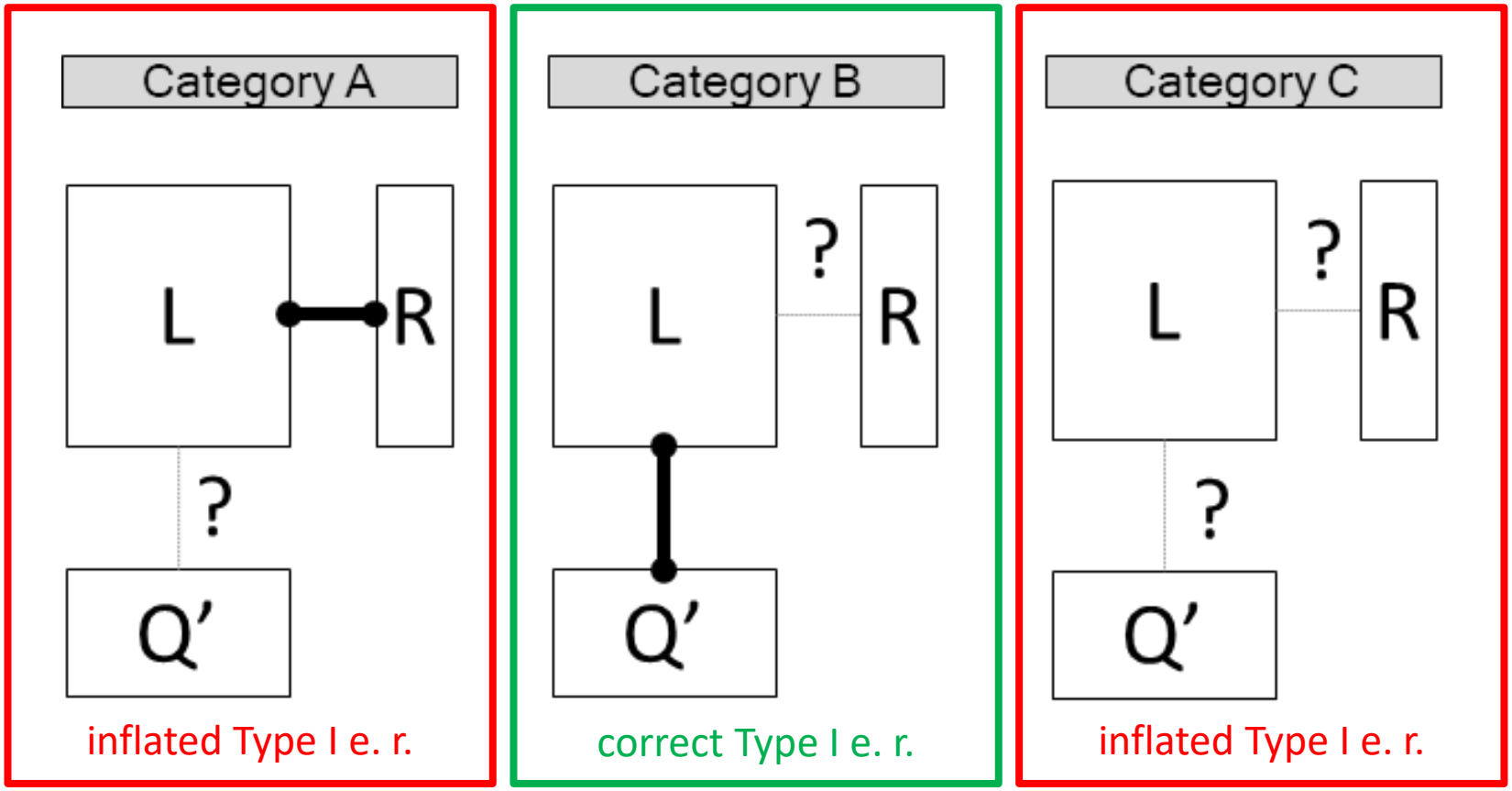
Proposal: introduce three categories of hypotheses and identify suitable test

Category C

- No explicit assumption
- e.g. relating traits to environment without explicit assumption whether traits are functional or not;
- or relating Ellenberg indicator values to environmental variables to test which environmental variable they most likely represent

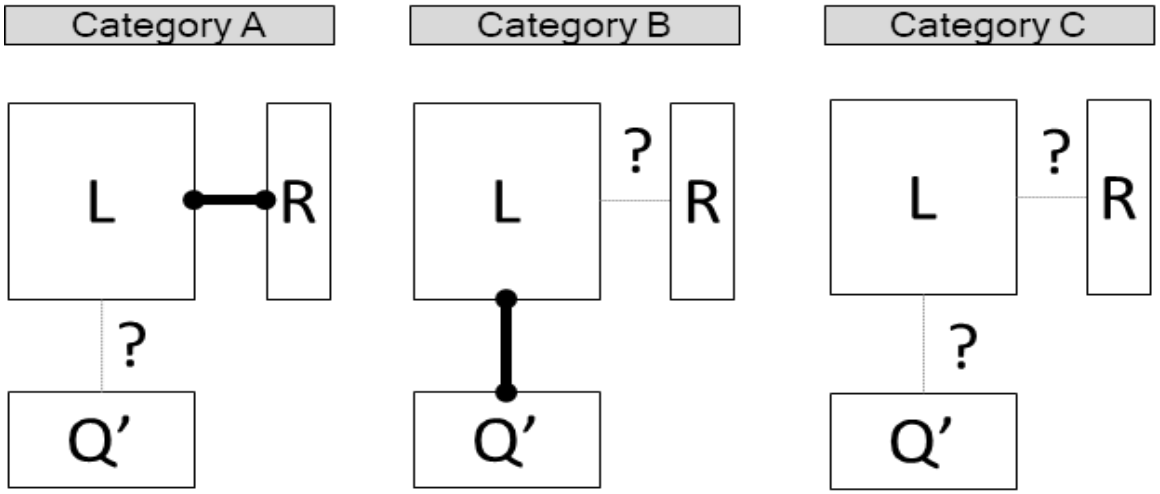
Standard test has inflated Type I error rate (too good to be true)

Proposal: introduce three categories of hypotheses and identify suitable test

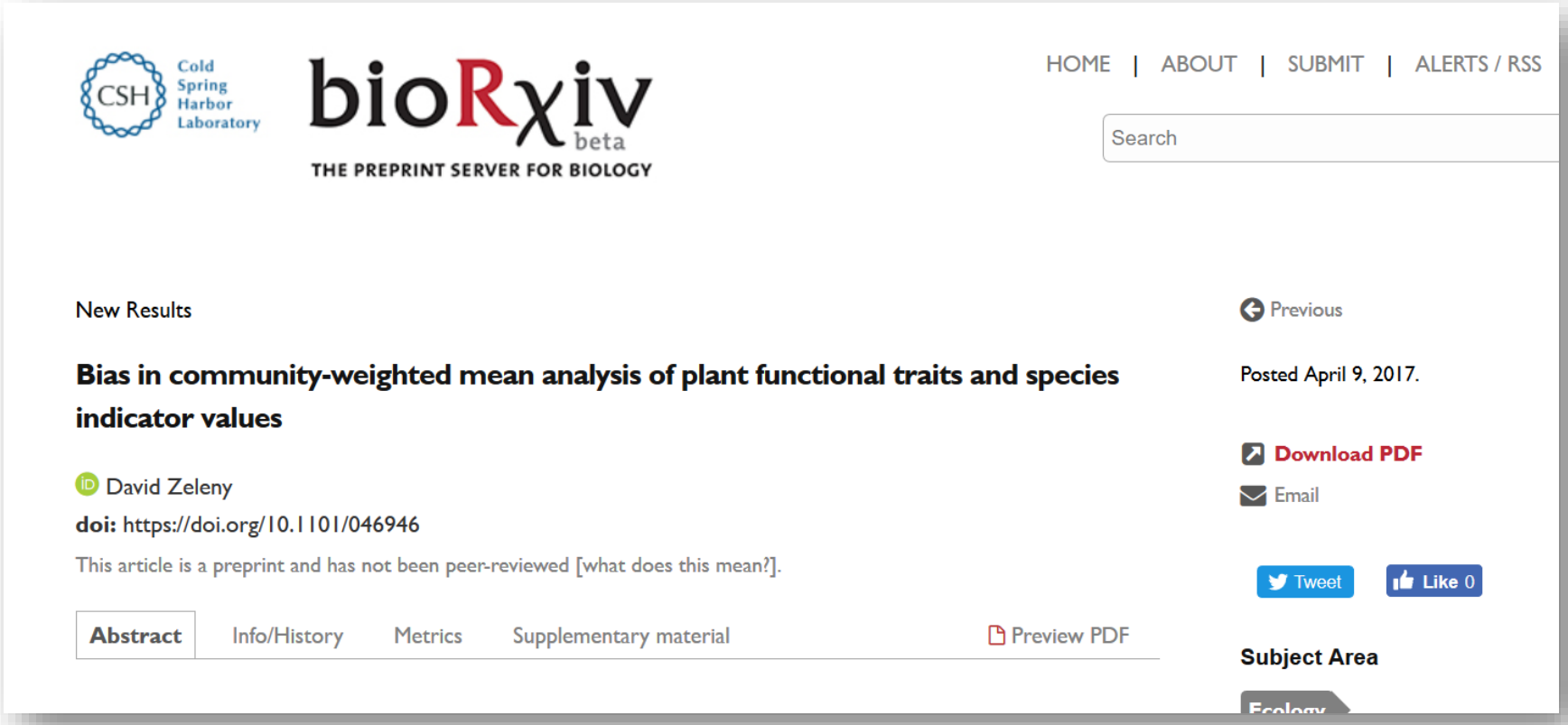


Explanation: ⊷ = an assumption of the relationship, **?** = what is being tested

Proposal: introduce three categories of hypotheses and identify suitable test



Recommended test:			
standard	NO	YES	NO
modified	YES	NO	NO
sequential with fourth-corner	YES	YES	YES



+

```
library (weimea)
```

Conclusions

- Standard test of CWM x sample attributes has inflated Type I error rate
- I propose to distinguish three categories of hypotheses and select the test accordingly – the sequential tests based on the fourth corner statistic may not always be necessary
- I suggest that the classification of hypotheses into categories may be an important tool also in interpreting the results of past studies
- Some issues remain to be solved: how general the approach based on the fourth corner statistic is, and how to deal with intrinsic sample attributes?

Thank you for your attention!

This study has been supported by
Ministry of Science and Technology, Taiwan
(MOST, 105-2621-B-002-00)