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

<table>
<thead>
<tr>
<th>Species</th>
<th>DCA1</th>
<th>DCA2</th>
<th>r²</th>
<th>Pr(&gt;r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>light</td>
<td>0.57495</td>
<td>0.81819</td>
<td>0.5187</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>temp</td>
<td>0.60487</td>
<td>0.79633</td>
<td>0.4351</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>cont</td>
<td>0.64357</td>
<td>0.76539</td>
<td>0.0194</td>
<td>0.429</td>
</tr>
<tr>
<td>moist</td>
<td>-0.88038</td>
<td>0.47428</td>
<td>0.8149</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>react</td>
<td>-0.79798</td>
<td>0.60269</td>
<td>0.3637</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>nutr</td>
<td>-0.98460</td>
<td>-0.17482</td>
<td>0.6727</td>
<td>0.001 ***</td>
</tr>
</tbody>
</table>

Species values randomized (by mistake)

~ 67% probability of result significant at P < 0.05
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; TWINSPLAN = two-way indicator species analysis

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) between species richness. Subsequently, a
CWM of specific leaf area (SLA) related to tree cover (Vltava valley data)

\[ r^2 = 0.123, \ P < 0.001 \]
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

**species attributes** – characteristics of individual species
- traits
- species indicator values
- species phylogenetic relationship
- species specialization values etc.

**sample attributes** – characteristics of individual samples (plots, localities)
- environmental variables
- experimental treatment
- group assignments (classification)
- diversity
- ordination scores

Intrinsic vs extrinsic sample or species 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
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

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

Slopes are given in units of trait/environment. \(P\)-values show results for testing the null hypothesis that slope = 0.
TESTING THE SPECIES TRAITS–ENVIRONMENT RELATIONSHIPS: THE FOURTH-CORNER PROBLEM REVISITED

STÉPHANE DRAY¹,³ AND PIERRE LEGENDRE²

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

CAJO J. F. TER BRAAK,¹,⁴ ANOUK CORMONT,² AND STÉPHANE DRAY³

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

STÉPHANE DRAY,¹,⁹ PHILIPPE CHOULER,²,³ SYLVAIN DOLÉDEC,⁴ PEDRO R. PERES-NETO,⁵ WILFRIED THUILLER,² SANDRINE PAVOINE,⁶,⁷ AND CAJO J. F. TER BRAAK⁸

¹Université de Lyon, F-69000, Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France
²Station Alpine J. Fourier, UMS CNRS-UJF 3370, Université J. Fourier, BP 53, F-38041 Grenoble, France
³Laboratoire d’Ecologie Alpine, UMR CNRS-UJF 5553, Université J. Fourier, BP 53, F-38041 Grenoble, France
⁴Université de Lyon, F-69000 Lyon, Université Lyon 1, CNRS, UMR5023, Laboratoire d’Ecologie des Hydrostèmes Naturels et Anthropisés, F-69622 Villeurbanne, France
⁵Canada Research Chair in Spatial Modelling and Biodiversity, Département des Sciences Biologiques, Université de Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, Quebec, H3C 3P8 Canada
⁶Laboratoire de Biologie des Populations, Université de Lyon, F-69000 Lyon, Université Lyon 1, CNRS, UMR6132, F-69622 Villeurbanne, France
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

Establishing trait-environment relationships in ecology is a major challenge. Community weighted means correlation (CWM) has emerged as an alternative method to the species niche centroid correlation (SNC) but it is not known whether the CWM method is valid in certain conditions. Using mathematical derivations and simulations, we show that the CWM method, when applied to the species niche centroid correlation, leads to inflated significance rates and unacceptably low statistical power even when the traits and the environment are highly correlated and the traits themselves are not linked. We show how the CWM method can be replaced by the Cheesell fourth-corner approach, which has higher statistical power. We propose an appropriate partitioning scheme to better test the trait-environment relationship and to improve the statistical power in detecting trait-environment associations. The approach is easily implementable in R and is applied for a contrast between methods and an application to a real-world data set.
Structural bias in aggregated species-level variables driven by repeated species co-occurrences: a pervasive problem in community and assemblage data

Bradford A. Hawkins¹,², Boris Leroy²,², Miguel Á. Rodríguez³, Alexander Singer⁴,⁵,⁶, Bruno Vilela³,², Fabricio Villalobos⁷,⁸, Xiangping Wang⁹ and David Zelený¹⁰

¹Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA, ²UMR 7208 BOREA, Muséum National d’Histoire Naturelle, Sorbonne Universités, Université Pierre et Marie Curie, Université de Caen, CNRS, IRD, Université des Antilles, Paris, France, ³Forest Ecology & Restoration Group, Department of Life Sciences, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain, ⁴German 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:

<table>
<thead>
<tr>
<th></th>
<th>Category A</th>
<th>Category B</th>
<th>Category C</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>modified</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>sequential with fourth-corner</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>
Bias in community-weighted mean analysis of plant functional traits and species indicator values

David Zeleny

doi: https://doi.org/10.1101/046946

This article is a preprint and has not been peer-reviewed [what does this mean?].

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)