

1 **Which results of the standard test in community weighted mean approach are**  
2 **too optimistic?**

3 David Zelený

4 ORCID: 0000-0001-5157-044X

5 Institute of Ecology and Evolutionary Biology, National Taiwan University, Roosevelt Rd. 1,

6 Taipei 10617, Taiwan

7 email: [zeleny@ntu.edu.tw](mailto:zeleny@ntu.edu.tw)

8 **# Running head:** Too optimistic CWM approach

9 **# Type of paper:** Synthesis

10 **# Page estimate:** ~9948 words (12.5 pages), tables (0.5 page), figures (1.5 page); **total ~ 14.5**

11 **pages**

12

## 13 **Abstract**

14 **Questions:** Community weighted mean (CWM) approach analyses the relationship species  
15 attributes (like traits or Ellenberg-type indicator values) to sample attributes (environmental  
16 variables). Recently it has been shown to suffer from inflated Type I error rate if tested by  
17 standard parametric or (row-based) permutation test. Results of many published studies are likely  
18 influenced, reporting overly optimistic relationships that are in fact merely a numerical artefact.  
19 Can we evaluate results of which studies are likely to be influenced and how much?

20 **Methods:** I suggest that hypotheses commonly tested by CWM approach are classified into three  
21 categories, which differ by assumption they make about the link of species composition to either  
22 species or sample attributes. I used a set of simulated and one simple real dataset to show how is  
23 the inflated Type I error rate influenced by data characteristics.

24 **Results:** For hypotheses assuming the link of species composition to species attributes, CWM  
25 approach with standard test returns correct Type I error rate. However, for the other two  
26 categories (assuming link of species composition to sample attributes or not assuming any link) it  
27 returns inflated Type I error rate and requires alternative tests to control for it (column-based and  
28 max test, respectively). Inflation index is negatively related to the beta diversity of species  
29 composition and positively to the strength of species composition-sample attributes relationship  
30 and the number of samples in the dataset. Inflation index is also influenced by modifying species  
31 composition matrix (by transformation or removal of species). The relationship of CWM with  
32 intrinsic species attributes is a case of spurious correlation and can be tested by column-based  
33 (modified) permutation test.

34 **Conclusions:** The concept of three hypothesis categories offers a simple tool to evaluate whether  
35 given study reports correct or inflated Type I error rate, and how inflated the rate can be.

36

37 **Keywords:** Ellenberg-type species indicator values; extrinsic attributes; fourth-corner approach;  
38 inflated Type I error rate; inflation index; intrinsic attributes; max test; modified permutation test;  
39 simulated data; species functional traits; species niche centroid approach

40

41 **Abbreviations:** CWM – community weighted mean; SNC – species niche centroid;

42

## 43 **Introduction**

44 A common task of community ecologists is relating species attributes to sample attributes using  
45 the matrix of species composition. Species attributes are characteristics of individual species, like  
46 species functional traits, ecological optima or phylogenetic age, while sample attributes are  
47 characteristics of individual samples, which can be either measured or estimated (environmental  
48 variables) or derived from the matrix of species composition itself (species richness, sample  
49 ordination scores). The matrix of species composition, which connects species and sample  
50 attributes, represents abundances (or presences-absences) of species in community samples. One  
51 way to find out whether there is a link between species and sample attributes is to calculate the  
52 mean of species attributes for species occurring in each sample weighted by relative species  
53 abundances (community-weighted mean, CWM), relate it to sample attributes, e.g. by correlation,  
54 and test this relationship by a relevant test. Here I call this method a *CWM approach* and use it as  
55 a general term including a wide range of analyses relating CWM of species attributes to sample  
56 attributes, where one or several CWMs and one or several sample attributes are involved. The  
57 mean of species attributes not weighted by species abundances is also included in CWM  
58 approach, since it is identical with CWM calculated on species composition matrix with species  
59 presence-absences instead of abundances.

60 In vegetation ecology, the two commonly used species attributes are plant traits and  
61 species indicator values. CWMs of plant traits can be related to environmental variables to  
62 demonstrate the effect of environmental filtering on trait-mediated community assembly (Díaz et  
63 al. 1998; Shipley 2010), or to predict changes in ecosystem properties (such as biomass

64 production or nutrient cycling; Garnier et al. 2004; Vile et al. 2006), or ecosystem services (like  
65 fodder production or maintenance of soil fertility; Díaz et al. 2007). CWMs of species indicator  
66 values, like those of Ellenberg et al. (1992) or Landolt (1977) for soil reaction, light, temperature  
67 and other factors, are used to estimate habitat conditions from known species composition of  
68 vegetation samples. These estimates are often related to soil, light or climatic variables  
69 (Ellenberg et al. 1992; Schaffers & Sýkora 2000), used for ecological interpretation of  
70 compositional changes in unconstrained ordination (Persson 1981) or ecological differences  
71 between groups of samples representing different vegetation types or treatments (Chytrý et al.  
72 2009). Other, more specific examples include relating the community specialization index to  
73 environmental variables (Clavero & Brotons 2010; Fajmonová et al. 2013; Carboni et al. 2016),  
74 or attempts to verify whether plant biomass can be estimated from tabulated plant heights and  
75 species composition as the mean of species heights weighted by their cover in a plot (Axmanová  
76 et al. 2012).

77         The CWM approach is also used in other fields, like biogeography (relating grid-based  
78 means of species properties, such as animal body size, to macroclimate or diversity; Hawkins &  
79 Diniz-Filho 2006), hydrobiology (relating trophic diatom index based on weighted mean of  
80 diatom indicator values to measured water quality parameters to assess its reliability; Kelly &  
81 Whitton 1995), or paleoecology (one of the transfer functions used to reconstruct acidification of  
82 lakes from fossil diatom assemblages preserved in lake sediments is based on weighted means of  
83 diatom optima along the pH gradient; ter Braak & Barendregt 1986; Birks et al. 1990).

84           The recent paper of Peres-Neto et al. (2017), focused on CWM approach, revealed  
85 several surprising facts. First, and perhaps the most important finding, is that standard tests  
86 analysing CWM-sample attributes relationship have inflated Type I error rate, returning more  
87 optimistic results than is warranted by the data. “Standard tests”, in the meaning used in current  
88 paper, include parametric tests like *t*-test for correlation and *F*-test for regression or ANOVA, or  
89 permutation tests randomising sample attributes (equivalent to randomising rows in the species  
90 composition matrix). Second, CWM correlation without applying weights has rather bad  
91 mathematical properties, and dominance of a single species can revert the sign of correlation  
92 even if the true trait differences are minimal (see also a dandelion example in Šmilauer & Lepš  
93 2014 and worked example in Appendix S1 of ter Braak et al. 2018). Third, the CWM approach is  
94 numerically related to the seemingly different fourth-corner problem (Legendre et al. 1997),  
95 which relates species and sample attributes via the species composition matrix without explicitly  
96 calculating weighted means of species attributes. Fourth, the ‘max test’ (Cormont et al. 2011),  
97 which solves the problem of inflated Type I error rate in the fourth-corner approach (ter Braak et  
98 al. 2012), does the same in the CWM approach. The max test undertakes two independent  
99 permutation tests, one testing species attributes-species composition link and the other sample  
100 attributes-species composition link, and chooses the higher P-value as a result. In conclusion,  
101 Peres-Neto et al. (2017) suggested to apply max test in all CWM analyses, and possibly replace  
102 the CWM approach with the more efficient fourth-corner approach.

103           Findings of Peres-Neto et al. (2017) will undoubtedly cause a revolution in the analysis of  
104 trait-environment, and generally, species attributes-sample attributes relationship; the max test  
105 should be included into the toolbox routinely used methods analysing trait-environment

106 relationship, and fourth-corner approach should be more attention than it had so far. It is also  
107 quite relevant to expect that scientific literature using CWM approach with standard tests is  
108 flooded by overly optimistic studies reporting significant relationships between various species  
109 and sample attributes which in fact are merely an analytical artefact. However, use of CWM  
110 approach has a long tradition in ecology, and quite often calculating CWM of species attributes  
111 and relating them to sample attributes is practical or required by theory. Many studies defining  
112 our current empirical knowledge about the trait-environment relationship or efficiency of  
113 Ellenberg-type indicator values have been published, and many studies will use this approach in  
114 future. What to do with that? How to recognise whether inflated Type error rate influences the  
115 results of certain study or not and if yes, how strong is the influence? Moreover, if CWM  
116 approach is used in future studies, is it always necessary to replace the standard tests by the max  
117 solution? These are some of the questions I will attempt to answer here.

118         First, I briefly review the use of CWM approach in vegetation ecology, its conceptual link  
119 to other methods analysing the relationship of sample and species attributes via a matrix of  
120 species composition, and current knowledge about the problem of inflated Type I error rate.  
121 Second, I suggest studies using CWM approach to be classified into one of the three categories,  
122 based on underlying assumptions about the link of species or sample attributes to species  
123 composition. Two of these categories return inflated Type I error rate in case that CWM  
124 approach is tested by the standard test, but only one of these categories requires the use of the  
125 max test as the only way to control for correct Type I error rate. Third, I acknowledge that  
126 sample attributes are of two types, extrinsic (measured independently of species composition  
127 matrix) and intrinsic (derived from species composition matrix), and discuss a special case of

128 CWM correlation with intrinsic species attributes. Finally, I use simulated community data to  
129 explore how is the rate of Type I error in standard CWM analysis influenced by data  
130 characteristics like beta diversity of species composition matrix, the strength of the link between  
131 sample attributes and species composition, and the number of samples in the dataset, and then  
132 show the same effect using a real vegetation dataset.

133

## 134 **Theory and Methods**

### 135 **CWM approach in the context of other methods**

136 Three objects are involved in the calculation of CWM approach: a vector of sample attributes ( $\mathbf{e}$ ,  
137 e.g. environmental variables), a matrix of species composition ( $\mathbf{L}$ , abundances or presences-  
138 absences of species in samples, with samples as rows and species as columns), and a vector of  
139 species attributes ( $\mathbf{t}$ , e.g. species traits); naming convention of variables follows Peres-Neto et al.

140 (2017). CWM of species attributes is calculated as  $c_i = \sum_{j=1}^S p_{ij} t_j$ , where  $S$  is the number of

141 species in a community,  $p_{ij}$  is the relative contribution of species  $j$  to the total abundance of  $i$ -th

142 sample, and  $t_j$  is the value of species attribute (“trait”) for the species  $j$ . Relative species

143 proportion  $p_{ij}$  can be calculated as  $l_{ij} / \sum_{j=1}^S l_{ij}$ , where  $l_{ij}$  is the abundance (or other measure,

144 such as biomass or presence-absence) of species  $j$  in the  $i$ -th sample and the denominator is the

145 sum of abundances for all species. The absolute values of  $p_{ij}$  (and consequently also the absolute

146 value of CWM) will be different if the denominator in the formula ( $\sum_{j=1}^S l_{ij}$ ) sums species



147 abundances across all species recorded in the data set, or only those for which the values of  
148 sample attribute ( $t_{ij}$ ) is available (and disregarding the others; see more in Discussion). Similar  
149 equations (with different notations) are reported in a number of studies, e.g. in Garnier et al.  
150 (2004) for CWM of species functional traits or in Diekmann (2003) for CWM of Ellenberg  
151 indicator values. CWM is either weighted or unweighted by (absolute or relative) species  
152 abundances, which is equivalent to saying that it is calculated on the matrix of species  
153 composition using raw (or relative) species abundances (weighted version) or presences-  
154 absences (unweighted version). Additionally, CWM can be also weighted by species amplitudes  
155 if these are available, where species with narrower amplitudes have a higher weight than species  
156 with broader amplitudes. This approach requires extending the CWM formula for amplitude  
157 parameter (Zelinka & Marvan 1961), and while commonly used in hydrology (e.g. diatom or  
158 saprophytic index, Kelly & Whitton 1995), it is rarely applied in vegetation ecology (but see  
159 Pepler-Lisbach 2008 using it for Ellenberg indicator values) and will not be further discussed  
160 here.

161 CWM is related to environmental variables (or other sample attributes) by a wide range  
162 of methods like correlation (called CWM correlation in the further text), regression or ANOVA,  
163 or more complex methods like mixed effect models, ordination (CWM-RDA and RLQ method,  
164 Kleyer et al. 2012; Dolédec et al. 1996) or correlation on distance matrices (Pillar et al. 2009).  
165 The strength and the direction of the relationship between CWM and sample attributes are  
166 quantified by relevant statistic (model parameters or effect size), and the significance of this  
167 statistic is often tested. The test is either parametric (e.g.  $t$ -test for correlation), or permutation

168 with the test statistic generated by reshuffling sample attributes (equivalent to permuting rows in  
169 species composition matrix, hence row-based permutation test, Fig. 1a).

170         Alternative methods analysing pairwise relationships of individual species attributes to  
171 individual sample attributes via the matrix of species composition include species niche centroids  
172 approach (SNC; ter Braak & Looman 1986) and the fourth-corner approach (Legendre et al.  
173 1997). While in the CWM approach sample attributes are related to the (weighted) mean of  
174 species attributes, in SNC approach the species attributes are related to the (weighted) mean of  
175 sample attributes (“niche centroids”). The fourth-corner approach (or the “fourth-corner  
176 problem”), in contrast, is not explicitly calculating weighted means of species or sample  
177 attributes, but combines all three objects ( $\mathbf{e}$ ,  $\mathbf{t}$ , and  $\mathbf{L}$ ) by inflating the original data tables (Dray  
178 & Legendre 2008). The original algorithm by Legendre et al. (1997) considered only presence-  
179 absence data in species composition matrix and introduced four different permutation tests, each  
180 aiming to test the different ecological hypothesis. Dray & Legendre (2008) extended the method  
181 also to quantitative species composition data and introduced universal two-step permutation test,  
182 later replaced by max test (ter Braak et al. 2012).

183         The fourth-corner problem, CWM and SNC approaches are in fact mathematically  
184 closely related (Peres-Neto et al. 2017). The fourth-corner statistic  $r$  is equal to the slope of the  
185 weighted linear regression between SNC of environmental variable and trait (Dray & Legendre  
186 2008) or CWM of trait and environmental variable (ter Braak et al. 2018), in the case that the  
187 regression is weighted, and traits with environmental variables are weighted standardized prior to  
188 calculation. Also, the weighted correlation of CWM of traits and environmental variable or SNC

189 of environmental variable and trait is related to the fourth-corner's  $r$ , and can be recalculated to  
190 each other using ratios between weighted standard deviations of CWM and traits, or SNC and  
191 environmental variable, respectively (equation 15 in Peres-Neto et al. 2017). Weights mentioned  
192 above are derived from the species composition matrix  $\mathbf{L}$ , as either total species abundances in  
193 samples (row sums in  $\mathbf{L}$ ) or sums of individual species abundances across all samples (column  
194 sums of  $\mathbf{L}$ ). Row sums of  $\mathbf{L}$  are used as weights in CWM regression and correlation and to  
195 weighted-standardise environmental variable in the regression, and the column sums of  $\mathbf{L}$  are  
196 used as weights in SNC regression and correlation and to weighted-standardise traits in  
197 regression. Note that two conceptually different types of weights are mentioned in the context of  
198 CWM (and also SNC) method. Abundances of individual species in individual samples (both  
199 CWM and SNC can also be calculated unweighted, equivalent to calculating them on matrix of  
200 species composition  $\mathbf{L}$  with species presences-absences instead of abundances), and plot or  
201 sample weights, calculated as row sums or column sums of species abundances in matrix  $\mathbf{L}$ ,  
202 respectively (and used as weights in weighted regression and correlation and to weighted  
203 standardize environmental variables and traits if necessary). Thus, if using the term “weighted”  
204 in case of CWM (or SNC) approach, it is advisable to specify which of the weights are meant.

205

### 206 **Inflated Type I error rate of standard test in CWM approach**

207 As mentioned above, Peres-Neto et al. (2017) showed that CWM approach might return overly  
208 optimistic results with inflated Type I error rate, falsely indicating the link between species and  
209 sample attributes where there is none. In fact, several previous studies indicated that testing

210 CWM-environment relationship is possibly problematic. Pillar et al. (2009) used column-based  
211 permutation test to assess link between CWM and environment using correlation between  
212 distance matrices in the study discriminating trait-convergence and divergence patterns in  
213 community assembly; they argue that “[t]he null model should retain most of the real data  
214 structures except for the one that is to be tested”. Jansen et al. (2011) calculated the relationship  
215 between mean Ellenberg indicator values or mean trait values and environmental variables and  
216 tested it by randomization test with permutation of species attributes, arguing that “[d]ue to the  
217 non-random co-occurrence of species in relevés... the correlation of mean trait values to  
218 measured site conditions can also be achieved by chance”. Zelený & Schaffers (2012) warned  
219 against the danger of overly optimistic results in the context of relating mean Ellenberg indicator  
220 values to ordination scores, assignment of samples into groups using cluster analysis, and species  
221 richness. They argued that these optimistic results are caused by CWM inheriting information  
222 about the compositional similarity between community samples and relating CWM to other  
223 variables having the same similarity issue causes the problem. They suggested that this  
224 relationship should either not be tested, or the “modified permutation test” with randomisation of  
225 species attributes should be used. Peres-Neto et al. (2012) discussed similar issue in the context  
226 of metacommunity phylogenetics, Šmilauer & Lepš (2014, p. 158) in the context of the CWM-  
227 RDA method and Hawkins et al. (2017) in the macroecological context when relating CWM of  
228 species traits to species richness.

229         Parallel to developments related to the CWM approach, Dray & Legendre (2008)  
230 identified the problem of inflated Type I error rate in the fourth corner (Legendre et al. 1997) if  
231 the fourth-corner statistic is tested by the row-based permutation method. Dray & Legendre

232 (2008) suggested the use of two-step testing procedure combining row- and column-based  
233 permutation tests together, a method which ter Braak et al. (2012) improved by introducing the  
234 sequential testing approach, called also max approach by later studies (ter Braak et al. 2017). The  
235 max test, first used by Cormont et al. (2011), is based on taking the maximum  $P$ -value from  
236 sequentially conducted row- and column-based permutation tests.

237         Hawkins et al. (2017) pointed out an important theoretical difference between *intrinsic*  
238 and *extrinsic* sample attributes, which differ in the relationship to the matrix of species  
239 composition. In general, *intrinsic* (sample or species) attributes are mathematically derived from  
240 the matrix of species composition, while *extrinsic* (sample or species) attributes have no  
241 mathematical relationship to it. Examples of intrinsic *sample* attributes include, e.g. species  
242 richness or diversity indices, sample ordination scores, sample assignments into clusters by  
243 numerical clustering, and also CWM calculated from species attributes and species composition;  
244 extrinsic sample attributes include measured or estimated environmental variables or grouping of  
245 samples according to experimental design. Intrinsic *species* attributes are also occasionally used  
246 (species scores on ordination axes or species optima calculated by weighted mean from species  
247 composition matrix) but are not further discussed here. The max test proposed by Peres-Neto et  
248 al. (2017) for CWM correlation applies to test the relationship between CWM and extrinsic  
249 sample attributes, and ter Braak et al. (2018) made it clear that there is no theoretical justification  
250 to use it for testing the relationship of CWM to intrinsic sample attributes. I suggest (in line with  
251 Zelený & Schaffers 2012) that test of CWM with intrinsic sample attributes can be done with  
252 column-based permutation test (modified permutation test sensu Zelený & Schaffers 2012), if we

253 consider such relationship as an example of spurious correlation (Brett 2004); more about this  
254 below (“*Spurious correlation*” of CWM with intrinsic sample attributes).

255

### 256 **Three categories of hypotheses tested by CWM approach**

257 I suggest that each hypothesis tested by the CWM approach fall into one of the three categories  
258 (labelled here as A, B or C, see Table 1 for a summary), depending on assumptions it makes  
259 about the link of matrix of species composition to either species attributes or sample attributes,  
260 respectively (Fig. 2). One may assume that the link exists if there is sufficient support for it  
261 either in the explicit formulation of a tested hypothesis or implicitly from the theoretical context  
262 of the study. The hypotheses in *category A* assume the link of species attributes to species  
263 composition ( $\mathbf{t} \leftrightarrow \mathbf{L}$ ), hypotheses in *category B* assume the link of sample attributes to species  
264 composition ( $\mathbf{e} \leftrightarrow \mathbf{L}$ ), and in hypotheses in *category C* does not assume any of the two links.

265       Indeed, the choice of the appropriate category may not always be straightforward. For  
266 example, trait studies testing whether the environment is filtering the species into a community  
267 via their functional traits routinely assume that such traits are functional and as such traits are  
268 considered to be linked to species composition (category A). This is reasonable in case that for  
269 studied trait there is sufficient evidence from other studies about its functional effect. However,  
270 this assumption may not be justified if the analysis is based on traits that are relatively easy to  
271 measure and thus readily available in databases, but which may not necessarily be the functional  
272 ones. Also, even the trait which is generally considered as functional does not need to be  
273 functional in the context of used dataset. Similarly, it may be reasonable to assume that species

274 composition is linked to sample attributes (**L-e**), e.g. if the study is based on experimental  
275 treatment which is known to change species composition and the question is focused on how  
276 these changes are reflected by sample attributes (e.g. Ellenberg-type indicator values, Chytrý et  
277 al. 2009).

278 As a simple rule to decide whether it is relevant to consider the existence of **L-t** or **L-e**  
279 link or not, one may ask whether it is interesting to test the existence of given link, or whether it  
280 would make sense to randomise species (**t**) or samples (**e**) attributes, respectively. If the answer  
281 is no, it may be safe to assume that given attributes are linked to species composition. In the case  
282 of functional traits example above, if there is a sufficient evidence to say that the trait is  
283 functional (e.g. experimental study, or previous empirical studies), it may be reasonable to  
284 assume that the link exists and does not need to be tested; if any doubt occurs, better to test it.  
285 The link between species composition and sample attributes (**L-e**) is tested by row-based test  
286 (parametric or permutation), the link between species composition and species attributes (**L-t**) by  
287 column-based permutation test, and both links simultaneously by the max test combining both  
288 row- and column-based tests together by selecting the larger *P*-value (Peres-Neto et al. 2017 and  
289 Fig 1b here). Even if max test seems to represent universal testing solution, in fact in categories  
290 A and B, the link which is assumed to exist does not need to be tested. This simplifies the test to  
291 either row-based (i.e. standard test), testing the link between sample attributes and species  
292 composition in category A, or column-based, testing the link between species attributes and  
293 species composition in category B. Only hypotheses in category C require both row- and  
294 column-based tests to be done, and max test was proved to control Type I error rate (Peres-Neto  
295 et al. 2017).

296 **Category A.** Studies in this category assume that species attributes are linked to species  
297 composition. For example, trait-based studies asking whether species traits can explain the effect  
298 of environmental filtering on species abundance in a community fall into this category. The null  
299 hypothesis, which states that sample attributes are not linked to species composition, can be  
300 tested by row-based (standard) parametric or permutation test.

301 **Category B.** Studies in this category assume that sample attributes are linked to species  
302 composition. Includes experimental studies in which the effect of experimental treatment  
303 (sample attribute) on species composition is acknowledged, and the question is about the  
304 response of species attributes to it. The null hypothesis, which states that species attributes are  
305 not linked to species composition, can be tested by column-based permutation test (also called  
306 modified permutation test in Zelený & Schaffers 2012).

307 **Category C.** Studies in this category assume no link between either species or sample attributes  
308 to species composition. Examples include empirical studies describing the general relationship  
309 between sample attributes and species attributes, without acknowledging any assumption based  
310 on the mechanism of such relationship (e.g. studies relating the CWM of traits to environmental  
311 variables without a priori assuming that traits are functional, allowing to question whether  
312 particular traits are linked to species composition or not). Studies with species indicator values  
313 relating mean indicator values to measured environmental variables also fit this category. To  
314 reject the null hypothesis, which states that there is no link between species or sample attributes  
315 and the matrix species composition, means to prove that both species and sample attributes are



316 linked to species composition, and this can be done by max test combining both row- and  
317 column-based tests.

318

### 319 **“Spurious correlation” of CWM with intrinsic sample attributes**

320 Examples of the relationship of CWM of species attributes with intrinsic sample attributes  
321 include analyses testing the relationship between CWM of Ellenberg-like indicator values and  
322 ordination scores (Zelený & Schaffers 2012) or CWM of traits with species richness (Hawkins et  
323 al. 2017). Wildi (2016) argued that testing such relationship violates the requirement on the  
324 independence of tested variables and should not be used, and ter Braak et al. (2018) warns  
325 against the use of the max test because it is not justified by the theory behind in this context. I  
326 suggest that this relationship can be considered as a case of “spurious correlations”, i.e. a  
327 relationship between compound variables calculated from the same parent variables (Pearson  
328 1897). Spurious correlations like  $X/Z \sim Y/Z$ ,  $X \sim Y/X$  or  $X+Y \sim Y$  (where X, Y and Z are  
329 variables related together) are ubiquitous in ecology, often encountered also in plant trait studies  
330 when one routinely calculates and tests the relationship between calculated traits (e.g. between  
331 specific leaf area, SLA, and leaf area, LA, where SLA is calculated as ratio between LA and leaf  
332 dry weight, LDW:  $SLA \sim LA = LA/LDW \sim LA$ ). Although opinions on how to deal with  
333 spurious correlations differ among researchers (Jackson & Somers 1991), general suggestion is  
334 to either avoid analysing relationship between compound variables, or to acknowledge their non-  
335 independence by testing their observed relationship (e.g. correlation) against the null expectation  
336 which would exist even if the parent variables are generated in random. For this, Jackson &

337 Somers (1991) and Brett (2004) suggested generating such null expectation by a permutation  
338 model, permuting the variable occurring only on one side of the equation.

339 CWM of species attributes and intrinsic sample attributes (like ordination scores or  
340 species richness) are both functions of a species composition matrix  $\mathbf{L}$ . CWM can be rewritten as  
341  $f_1(\mathbf{t}, \mathbf{L})$  and intrinsic samples attributes as  $f_2(\mathbf{L})$ , where  $\mathbf{t}$  is the vector of species attributes  
342 (Ellenberg-type indicator values, traits) and  $\mathbf{L}$  is the matrix of species composition. In the  
343 relationship  $f_1(\mathbf{t}, \mathbf{L}) \sim f_2(\mathbf{L})$ , the compositional matrix  $\mathbf{L}$  is a parent variable occurring on both  
344 sides of the equation, in the same sense as in the spurious correlation. The null expectation of the  
345 test statistic can be calculated by permuting the trait values among species as in the modified  
346 permutation test suggested by Zelený & Schaffers (2012). Modified test changes the original null  
347 hypothesis of *no relationship between CWM of species attributes and intrinsic sample attributes*  
348 (i.e.  $f_1(\mathbf{t}, \mathbf{L}) \not\sim f_2(\mathbf{L})$ ) into *no relationship between species attributes and intrinsic sample*  
349 *attributes* (i.e.  $\mathbf{t} \not\sim f_2(\mathbf{L})$ ). In this way, the modified permutation test remains a valid tool to  
350 correct for inflated Type I error rate when relating CWM of species attributes to sample  
351 attributes (e.g. relating mean Ellenberg-type indicator values to scores of unconstrained  
352 ordination, Zelený & Schaffers 2012).

353

### 354 **Dependence of inflated Type I error rate on data characteristics**

355 In this section, I illustrate how is the inflation of Type I error rate in CWM approach dependent  
356 on three dataset characteristics: compositional heterogeneity (beta diversity) of the species  
357 composition matrix, the strength of the link between sample attributes and species composition

358 (L-e link), and the number of samples in the community matrix. For this, I use CWM correlation  
359 with standard parametric test and apply it on a number of simulated community datasets. Then, I  
360 use the real vegetation dataset with Ellenberg-type indicator values to show how the inflation  
361 depends on the strength of the environment-species composition relationship.

### 362 *Design of the simulation study*

363 The algorithm generating simulated community data is an extension of COMPAS model  
364 proposed by Minchin (1987). Here, I used model structured by two virtual ecological gradients,  
365 which is an extension of one-gradient implementation by Fridley et al. (2007). Along each  
366 gradient, a number of unimodal species response curves was generated, where each response  
367 curve quantifies the probability with which an individual found in a given gradient location is  
368 assigned to given species. Species composition of individual samples was then generated by  
369 randomly selecting locations along both gradients and assigning given number of individuals  
370 (100) into species according to species probabilities at given gradient location. The first gradient  
371 is used to define sample and species attributes (locations of samples equals to sample  
372 ‘environmental variable’, while optima of species response curve equals to species ‘trait’), while  
373 the second gradient is used to modify the beta diversity of the whole dataset (increasing its length  
374 together with proportional increase in the number of species results in increased beta diversity of  
375 the species composition matrix). Species niche widths are generated as random numbers of  
376 uniform distribution between 500 and 1000 units, independently for each gradient. The effective  
377 length of the first gradient was arbitrarily set to 500 units (the true length is 1500 units, but only  
378 the range between 500 to 1000 units is populated by samples, to avoid gradient edges with a

379 lower density of response curves). The effective length of the second gradient varied between  
380 500 to 4500 units (also with extra 500 units at each side). As a result, each simulated community  
381 data set includes a matrix of sample attributes ('environmental variable',  $\mathbf{e}$ ), species composition  
382 ( $\mathbf{L}$ ) and species attributes ('traits',  $\mathbf{t}$ ), where sample attributes and species attributes are linked to  
383 species composition. Because the aim is to show what is the probability that CWM correlation  
384 will be significant even if the null hypothesis is true (i.e. species attributes are not related to  
385 species composition), species attributes were permuted to remove their link to species  
386 composition.

387       Beta diversity of the dataset was modified by increasing the length of the second gradient.  
388 I assumed that 500 units of the second gradient represent one community, i.e. enlarging the  
389 second gradient from 500 to 5000 units (by steps of 500 units) generated datasets of increasing  
390 beta diversity (with 1 to 9 communities). A dataset with a maximum number of communities was  
391 also included (max), in which the data are reshaped in the way that no species are shared among  
392 any pair of samples. The strength of the relationship between species composition and sample  
393 attributes ( $\mathbf{L}$ - $\mathbf{e}$  strength) was manipulated by adding random noise to generated values of sample  
394 attributes  $\mathbf{e}$ . I also included one intrinsic sample attribute, mathematically derived from the  
395 matrix of species composition ( $\mathbf{L}$ ) by an unconstrained ordination (sample scores along the first  
396 axis of correspondence analysis calculated on log-transformed species composition data, denoted  
397 as CA1). The number of samples was manipulated by increasing the density of locations along  
398 both gradients where communities were generated, while keeping the length of the gradients  
399 constant; this mimics the real situation of sampling increasing number of community samples  
400 within the same range of compositional heterogeneity.

401 I prepared two scenarios, each with one fixed and two variable data characteristics. In  
402 *scenario 1*, the number of samples was fixed (100 samples), while beta diversity and the strength  
403 of **L-e** relationship varied; for each combination of beta diversity (1, 3, 5, 7, 9 and max.  
404 communities) and **L-e** strength (0.0, 0.2, 0.4, 0.6, 0.8, 1 and CA1) I generated 1000 datasets. In  
405 *scenario 2*, the **L-e** strength was kept fixed (0.6), while the number of samples and beta diversity  
406 varied; for each combination of sample size ( $25 \times 2^n$  samples with  $n = \{0, 1, 2, 3, 4, 5\}$ ) and the  
407 same levels of beta diversities as in scenario 1) I generated 1000 datasets. For each dataset, I  
408 related CWM of **t** (weighted by species abundances) with **e** by Pearson's  $r$  correlation and tested  
409 the significance by standard parametric  $t$ -test, and then by permutation max test (199  
410 permutations, with absolute  $t$ -value as a test statistic). I quantified the inflation of Type I error  
411 rate in CWM correlation for each combination of data characteristics by inflation index (sensu  
412 Lennon 2000) calculated as  $I(\alpha) = N_o/N_e$ , where  $\alpha$  is the nominal significance level,  $N_o$  is the  
413 number of 'observed' correlations significant at  $\alpha$  level, and  $N_e$  is the number of 'expected'  
414 correlation significant at  $\alpha$  level (calculated as  $N_e = \alpha N_{total}$ , where  $N_{total}$  is the total number of  
415 tests). Inflation index quantifies how many times more likely we are to find significant result  
416 compared to the test with correct Type I error rate; test with inflation index close to unity has  
417 correct Type I error rate. I plotted the inflation index  $I(\alpha = 0.05)$  against beta diversity and the  
418 strength of **L-e** link (scenario 1) or a number of samples and beta diversity (scenario 2).

#### 419 *Design of real data study*

420 Example dataset of real data contain forest vegetation plots sampled by me on the slopes of deep  
421 valley of river Vltava, Czech Republic (Zelený & Chytrý 2007). The total of 97 plots of  $10 \times 15$  m  
422 were sampled at even distances along the transect running along the valley slope. All vascular

423 plant species were recorded and their cover estimated using Braun-Blanquet scale (Westhoff &  
424 van der Maarel 1978). A subset of 11 environmental variables measured or estimated for each  
425 plot was selected (details in Zelený & Chytrý 2007). Species attributes used in this analysis are  
426 Czech indicator values for light, temperature, moisture, reaction and nutrients, which are  
427 Ellenberg-type species indicator values recalibrated for territory of the Czech Republic (Chytrý  
428 et al. 2018). For the analysis presented here, species composition data include only herbs  
429 sampled in the forest understory, since indicator values for light are defined only for herbs and  
430 juveniles of woody species. Only species that have all five indicator values available were  
431 included, others were removed from both species composition matrix and matrix of indicator  
432 values; this is to guarantee that all calculated CWM are based on species composition datasets  
433 with identical beta diversity. Additionally to 11 extrinsic (measured or estimated) environmental  
434 variables, I included also one intrinsic variable, scores of samples along the first axis of  
435 correspondence analysis calculated on (log transformed) species composition data. As a result,  
436 three matrices were used for CWM correlation: environmental variables (97 samples  $\times$  12  
437 variables), species composition (97 samples  $\times$  103 species) and Czech indicator values (103  
438 species  $\times$  5 indicator values). CWM was calculated as species mean weighted by estimated  
439 species abundances transformed into the percentage scale. The strength of **L-e** link for each  
440 environmental variable was quantified as variance ( $R^2_{CCA}$ ) this variable explains in canonical  
441 correspondence analysis (CCA) on log-transformed species composition data, rescaled to  
442 maximum variance one explanatory variable could theoretically explain (equal to variance  
443 represented by the first axis of correspondence analysis calculated on same species composition  
444 data; Šmilauer & Lepš 2014). The use of CCA is inspired by link between its unconstrained

445 version (CA) and fourth-corner approach (Peres-Neto et al. 2017), and although the ordination  
446 based on chi-square distances may not be the best method for CWM correlation (which does not  
447 apply species and sample weights in calculation), I use it here as a reasonable proxy. Note that  
448 the strength of **L-e** in analysis of simulated and real data is quantified by different methods; in  
449 simulated data, the strength is manipulated *post hoc* by adding certain level of noise to values of  
450 sample attributes (which by construction has strong **L-e** link), while in real data, the strength is  
451 calculated as variance in **L** explained by **e**. Since in real study the species composition data are  
452 the same for each combination of **t** and **e**, beta diversity is fixed and  $R^2_{CCA}$  reflects only the **L-e**  
453 strength. If the beta diversity was left to vary (e.g. by calculating CWM for species attributes  
454 with missing values of **t**), the  $R^2_{CCA}$  would reflect both **L-e** strength and beta diversity of **L**.

455 All analyses were done in R-project (version 3.5.0, R Foundation for Statistical  
456 Computing, Vienna, Austria, <https://www.R-project.org/>); complete R-script is available in  
457 Appendix S1. Simulated data were generated by package *simcom* (Zelený, unpublished), and  
458 CWM correlation was calculated by *weimea* (Zelený, unpublished; source code of v. 0.1.10 in  
459 Appendix S2, the latest version at <https://github.com/zdealveindy/weimea>).

460

## 461 **Results**

462 In an analysis based on simulated data, all three data characteristics (beta diversity, the strength  
463 of **L-e** link and sample size) influenced the inflation index of CWM correlation tested by  
464 standard parametric test (Fig. 3). The inflation index is negatively related to beta diversity and  
465 positively to the strength of **L-e** link (Fig. 3a,b). In the case of maximum beta diversity (samples

466 does not share any species) the inflation index is approaching unity for all strengths of **L-e** link.  
467 Inflation index is also positively related to the number of samples (Fig. 3c), with the highest  
468 inflation index for the most homogeneous community (number of communities = 1); for the most  
469 heterogeneous community (maximum beta diversity) the inflation index oscillates around unity.  
470 The max test applied on the same data removes the problem of inflated Type I error rate from all  
471 combinations of the three data characteristics (returning inflation index close to 1).

472 In the analysis of real data, beta diversity of the species composition data and the number  
473 of samples were fixed, and only the strength of **L-e** varied (differ among individual  
474 environmental variables). Those environmental variables with a stronger link to species  
475 composition were significantly ( $P < 0.05$ ) related to higher number of CWM of indicator values,  
476 with intrinsic variable (CA1) related to all five (Fig. 4a). In an analysis where randomly  
477 generated ones replaced real indicator values, the inflation index increased with the strength of  
478 **L-e** relationship, with values over 8 for environmental variables most strongly related to  
479 environment and almost 10 for CA1 (Fig. 4b).

480

## 481 **Discussion**

482 To avoid inflated Type I error rate in CWM approach, Peres-Neto et al. (2017) suggested using  
483 the max test as a universal solution. I suggest that as an alternative to this “one-fits-all” solution,  
484 it is useful to fully clarify what are the underlying assumptions the analysed question is putting  
485 on the links between members in the game, namely links of species composition to species  
486 attributes (**L-t**) or sample attributes (**L-e**), respectively. Standard (row-based parametric or



487 permutation) test controls for Type I error rate for the hypothesis in category A, and results of  
488 these studies, therefore, do not need to be considered as overly optimistic. In contrast, hypotheses  
489 in categories B and C requires alternative testing approach, namely column-based (modified)  
490 permutation test (B) and max test (C), to control for the Type I error rate. This concept can be  
491 useful for published studies using CWM approach with standard test, for which one can either  
492 clarify whether in the context of given study (with explicitly formulated hypothesis) the standard  
493 test returns correct Type I error rate (category A) and if not, whether it is possible to formulate  
494 an alternative hypothesis for which the presented results would be valid.

495         If the study fits the category for which Type I error rate of standard tests is inflated  
496 (category B or C), one can evaluate what is the probability that the reported values are overly  
497 optimistic. For this, information about beta diversity, the strength of **L-e** relationship and the  
498 sample size is needed, or needs to be calculated from the original data (if available). This can  
499 also help with conducting a meta-analysis in the future which would evaluate the scale of the  
500 problem (how many published studies report overly optimistic results). Indeed, in many  
501 published studies the data characteristics are not reported, and original data are not available;  
502 then only a rough guess whether the risk is high or low is possible based on available data  
503 description. Such guesses are, indeed, only approximate, and re-analysis using the original data  
504 is needed to get an exact answer.

505         Transformation of species abundances (e.g. square-root, log, or presence-absence) will  
506 influence the beta diversity of species composition data (and possibly of the strength of **L-e** link)  
507 and consequently also inflation index in CWM approach. For example, one may ask whether

508 traits or indicator values are better related to the environment if raw abundances or presences-  
509 absences are used for CWM calculation (Hill & Carey 1997; Pakeman et al. 2007). In this sense,  
510 the practice differs between the use of traits and indicator values. For traits, the weighting of  
511 individual species values by species abundances in the community is justified by Grime's Mass  
512 ratio hypothesis, which states that the functional effect of given species is proportional to its  
513 relative contribution to the total biomass of the community (Grime 1998). In contrast, CWM of  
514 Ellenberg-type species indicator values more often calculated unweighted by species abundances  
515 (i.e. calculated from presence-absence species composition data), because even species with low  
516 abundance or biomass can be a good indicator of environmental conditions (Ellenberg et al.  
517 1992). Attempts to answer whether raw abundances or transformed data should be used to  
518 calculate CWM are usually done by calculating both CWM of raw and transformed species  
519 attributes using the same dataset and relating them to the same sample attributes, including  
520 testing the significance by standard tests (Pakeman et al. 2007). This approach, however, does  
521 not allow separating the conceptual effect of species data transformation from a mere artefact  
522 caused by the fact that data transformation influences inflation index by changes in data  
523 characteristics.

524 Inflation index of standard test in CWM approach is likely to be also influenced by  
525 removal of species from species composition matrix, which changes the beta diversity of species  
526 composition and strength of L-e relationship. Species are usually removed because they are  
527 missing value for given species attribute (e.g. traits measured only for a subset of dominant  
528 species, or indicator values without assigning values to generalists) or because of some arbitrary  
529 decision (e.g. removing rare species). If more species attributes are related by CWM approach to

530 the same sample attribute using the same species composition matrix, and if these attributes have  
531 different proportion and identity of missing species, resulting inflation index can differ among  
532 species attributes. This can also bias results of studies that explicitly ask about the sensitivity of  
533 CWM approach to missing species values (Ewald 2003; Pakeman & Quested 2007) if these are  
534 based on comparing the number of significant relationships of CWM between the same species  
535 and sample attributes on datasets with increasing proportion of removed species.

536       Species with missing values of species attributes values are treated differently in CWM  
537 approach applied on traits and Ellenberg-type indicator values. For traits, CWM is often defined  
538 as the mean weighted by relative contributions of species into overall biomass, where overall  
539 biomass also includes species for which traits were not measured (e.g. Garnier et al. 2004), while  
540 for indicator values, species without indicator values (often generalists) are simply disregarded  
541 from the calculation (Diekmann 2003). For traits, this is equivalent to calculating species relative  
542 contribution  $p_{ij}$  from absolute species abundance divided by abundance sum of all species present  
543 in the community (including those with missing trait values). For indicator values, in contrast,  
544 species with missing indicator values are first removed from  $\mathbf{L}$  and  $p_{ij}$  is calculated as  $l_{ij}$  divided  
545 by sum of  $l_{ij}$  for species left in the matrix. Considering or disregarding the species without  
546 species attributes in the CWM calculation changes the absolute value of CWM and in result also  
547 the parameters estimated and tested by CWM approach.

548       In this study, I explicitly ignored intraspecific variation in species attributes, using only  
549 dataset-wide mean species attribute values. Indeed, intraspecific variation is important, both in  
550 the context of functional traits (Albert et al. 2012) and potentially also Ellenberg-type indicator

551 values (Peppler-Lisbach 2008). In case of traits, intraspecific variability can be considered by  
552 calculating CWM values from the site- or treatment-specific species trait values (Lepš et al.  
553 2011). In the case of Ellenberg-type indicator values, species ecological amplitude can be  
554 implemented as extra weight in CWM formula (Peppler-Lisbach 2008). Whether and how much  
555 is the Type I error rate of such calculations inflated remain to be tested.

556 Finally, the relevant consideration is whether the CWM approach is the best analytical  
557 solution for the question we aim to answer. In cases when the question is explicitly focused on  
558 relating community-level values of species attributes, like mean Ellenberg-like species indicator  
559 values (serving as an estimate of ecological conditions for individual sites) or the CWM of traits  
560 (as one of the functional-diversity metrics and as a community-level trait value) the use of CWM  
561 approach is entirely justified. In other cases, when the question is focused on relating individual  
562 species-attributes to sample attributes, the fourth corner approach should be considered as it is  
563 more powerful in detecting the sample attribute-species attribute relationship (Peres-Neto et al.  
564 2017).

565

## 566 **Conclusions**

567 The CWM approach with standard (row-based) test returns correct Type I error rate only in case  
568 the tested hypothesis assume that species composition is linked to species attributes. In other  
569 cases, the Type I error rate of the standard test is inflated, and the inflation index depends on the  
570 interaction between the beta diversity of species composition matrix, the strength of the  
571 relationship between species composition and sample attributes, and the number of samples in

572 the analysis. An alternative to standard test is a column-based or max test, respectively,  
573 controlling Type I error rate if species composition is linked to sample attributes (column-based  
574 test) or no link is assumed (max test). This concept can be used to evaluate whether results of  
575 studies using CWM approach with standard test report correct or inflated Type I error rate, and if  
576 inflated, how much.

577

## 578 **Acknowledgements**

579 My thanks go to Cajo ter Braak and two anonymous reviewers for critical comments on previous  
580 versions of this manuscript. This study was supported by the Ministry of Science and  
581 Technology, Taiwan (106-2621-B-002-003-MY3).

582

## 583 **Supporting information**

584 **Appendix S1.** R-code used to calculate simulated data and real data analysis.

585 **Appendix S2.** Source code for the R library *weimea*, version 0.1.10 (latest version can be  
586 installed from <https://github.com/zdealveindy/weimea/>).

587

## 588 **References**

- 589 Albert, C. H., de Bello, F., Lavorel, S. & Thuiller, W. 2012. On the importance of intraspecific  
590 variability for the quantification of functional diversity. *Oikos* 121: 116–126.
- 591 Axmanová, I., Tichý, L., Fajmonová, Z., Hájková, P., Hettenbergerová, E., Li, C.-F., Merunková,  
592 K., Nejezchlebová, M., Otýpková, Z., Vymazalová, M. & Zelený, D. 2012. Estimation of  
593 herbaceous biomass from species composition and cover. *Applied Vegetation Science* 15: 580–  
594 589.
- 595 Birks, H. J. B., Line, J.M., Juggins, S., Stevenson, A.C. & ter Braak, C.J.F. 1990. Diatoms and  
596 pH reconstruction. *Philosophical Transactions of the Royal Society B Biological Sciences*  
597 327: 263–278.
- 598 Brett M.T. 2004. When is a correlation between non-independent variables “spurious”? *Oikos*  
599 105: 647–656.
- 600 Carboni, M., Zelený, D. & Acosta, A. 2016. Measuring ecological specialization along a natural  
601 stress gradient using a set of complementary niche breadth indices. *Journal of Vegetation*  
602 *Science* 27: 892–203.
- 603 Chytrý, M., Hejčman, M., Hennekens, S.M. & Schellberg, J. 2009. Changes in vegetation types  
604 and Ellenberg indicator values after 65 years of fertilizer application in the Rengen Grassland  
605 Experiment, Germany. *Applied Vegetation Science* 12: 167–176.
- 606 Chytrý, M., Tichý, L., Dřevojan, P., Sádlo, J. & Zelený, D. 2018. Ellenberg-type indicator values  
607 for the Czech flora. *Preslia* 90: 83–103.

- 608 Clavero, M. & Brotons, L. 2010. Functional homogenization of bird communities along habitat  
609 gradients: accounting for niche multidimensionality. *Global Ecology and Biogeography* 19: 684–  
610 696.
- 611 Cormont, A., Vos, C.C., van Turnhout, C.A.M., Foppen, R.P.B. & ter Braak, C.J.F. 2011. Using  
612 life-history traits to explain bird population responses to changing weather variability. *Climate*  
613 *Research* 49: 59–71.
- 614 Díaz, S., Cabido, M. & Casanoves, F. 1998. Plant functional traits and environmental filters at a  
615 regional scale. *Journal of Vegetation Science* 9: 113–122.
- 616 Díaz, S., Lavorel, S., de Bello, F., Quétler, F., Grigulis, K. & Robson, T.M. 2007. Incorporating  
617 plant functional diversity effects in ecosystem service assessments. *Proceedings of the National*  
618 *Academy of Sciences USA* 104: 20684–20689.
- 619 Diekmann, M. 2003. Species indicator values as an important tool in applied plant ecology – a  
620 review. *Basic and Applied Ecology* 4: 493–506.
- 621 Dolédec, S., Chessel, D., ter Braak, C.J.F. & Champely, S. 1996. Matching species traits to  
622 environmental variables: a new three-table ordination method. *Environmental and Ecological*  
623 *Statistics* 3: 143–166.
- 624 Dray, S. & Legendre, P. 2008. Testing the species traits-environment relationships: the fourth-  
625 corner problem revisited. *Ecology* 89: 3400–3412.

- 626 Ellenberg, H., Weber, H.E., Düll, R., Wirth, V., Werner, W. & Paulissen, D. 1992. Zeigerwerte  
627 von Pflanzen in Mitteleuropa. Second Edition. *Scripta Geobotanica* 18: 1–248.
- 628 Ewald, J. 2003. The sensitivity of Ellenberg indicator values to the completeness of vegetation  
629 relevés. *Basic and Applied Ecology* 4: 507–513.
- 630 Fajmonová, Z., Zelený, D., Syrovátka, V., Vončina, G. & Hájek, M. 2013. Distribution of habitat  
631 specialists in semi-natural grasslands. *Journal of Vegetation Science* 24: 616–627.
- 632 Fridley, J.D., Vandermast, D.B., Kuppinger, D.M., Manthey, M. & Peet, R.K. 2007. Co-  
633 occurrence based assessment of habitat generalists and specialists: a new approach for the  
634 measurement of niche width. *Journal of Ecology* 95: 707–722.
- 635 Garnier, E., Cortez, J., Billès, G., Navas, M.L., Roumet, C., Debussche, M., Laurent, G.,  
636 Blanchard, A., Aubry, D., (...) & Neill, C. 2004. Plant functional markers capture ecosystem  
637 properties during secondary succession. *Ecology* 85: 2630–2637.
- 638 Grime, J.P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects.  
639 *Journal of Ecology* 86: 902–910.
- 640 Hawkins, B.A. & Diniz-Filho, J.A.F. 2006. Beyond Rapoport's rule: evaluating range size  
641 patterns of New World birds in a two-dimensional framework. *Global Ecology and*  
642 *Biogeography* 15: 461–469.
- 643 Hawkins, B.A., Leroy, B., Rodríguez, M.Á., Singer, A., Vilela, B., Villalobos, F., Wang X. &  
644 Zelený D. 2017. Structural bias in aggregated species-level variables driven by repeated species



- 645 co-occurrences: a pervasive problem in community and assemblage data. *Journal of*  
646 *Biogeography* 44: 1199–1211.
- 647 Hill, M.O. & Carey, P.D. 1997. Prediction of yield in the Rothamsted Park Grass Experiment by  
648 Ellenberg indicator values. *Journal of Vegetation Science* 8: 579-586
- 649 Jackson, D.A. & Somers, K.M. 1991. The spectre of ‘spurious’ correlations. *Oecologia* 86: 147–  
650 151.
- 651 Jansen, F., Ewald, J. & Zerbe, S. 2011. Ecological preferences of alien plant species in North-  
652 Eastern Germany. *Biological Invasions* 13: 2691–2701.
- 653 Kelly, M.G. & Whitton, B.A. 1995. Biological monitoring of eutrophication in rivers.  
654 *Hydrobiologia* 384: 55–67.
- 655 Kleyer, M., Dray, S., de Bello, F., Lepš, J., Pakeman, R.J., Strauss, B., Thuiller, W. & Lavorel, S.  
656 2012. Assessing species and community functional responses to environmental gradients: which  
657 multivariate methods? *Journal of Vegetation Science* 23: 805-821.
- 658 Landolt, E. 1977. Ökologische Zeigerwerte zur Schweizer Flora. *Veröffentlichungen des*  
659 *Geobotanischen Institutes der Eidgenössischen Technischen Hochschule, Stiftung Rübel, Zürich*  
660 64: 1–208.
- 661 Legendre, P., Galzin, R. & Harmelin-Vivien, M.L. 1997. Relating behavior to habitat: solutions  
662 to the fourth-corner problem. *Ecology* 78: 547–562.
- 663 Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* 23: 101–113.

- 664 Lepš, J., de Bello, F., Šmilauer, P. & Doležal, J. 2011. Community trait response to environment:  
665 disentangling species turnover vs intraspecific trait variability effects. *Ecography* 34: 856–863.
- 666 Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a  
667 comprehensive model. *Vegetatio* 71: 145–156.
- 668 Pakeman, J. & Quested, H.M. 2007. Sampling plant functional traits: What proportion of the  
669 species need to be measured? *Applied Vegetation Science* 10: 91–96.
- 670 Pakeman, R.J., Garnier, E., Lavorel, S., Ansquer, P., Castro, H., Cruz, P., Doležal, J., Eriksson,  
671 O., Freitas, H., ..., & Vile, D. 2008. Impact of abundance weighing on the response of seed traits  
672 to climate and land use change. *Journal of Ecology* 96: 355–366.
- 673 Pearson, K. 1897. On a form of spurious correlation which may arise when indices are used in  
674 the measurement of organs. *Proceedings of the Royal Society of London* 60: 489–498.
- 675 Peppler-Lisbach C. 2008. Using species-environment amplitudes to predict pH values from  
676 vegetation. *Journal of Vegetation Science* 19: 437–444.
- 677 Peres-Neto, P.R., Dray, S. & ter Braak, C.J.F. 2017. Linking trait variation to the environment:  
678 critical issues with community-weighted mean correlation resolved by the fourth-corner  
679 approach. *Ecography* 40: 806–816.
- 680 Peres-Neto, P.R., Leibold, M.A. & Dray, S. 2012. Assessing the effects of spatial contingency  
681 and environmental filtering on metacommunity phylogenetics. *Ecology* 93: S14–S30.

- 682 Persson S. 1981. Ecological indicator values as an aid in interpretation of ordination diagrams.  
683 *Journal of Ecology* 69: 71–84.
- 684 Pillar, V.D., Duarte, L.S., Sosinski, E.E. & Joner, F. 2009. Discriminating trait-convergence and  
685 trait-divergence assembly patterns in ecological community gradients. *Journal of Vegetation*  
686 *Science* 20: 334–348.
- 687 Schaffers, A.P. & Sýkora, K.V. 2000. Reliability of Ellenberg indicator values for moisture,  
688 nitrogen and soil reaction: comparison with field measurements. *Journal of Vegetation Science*  
689 11: 225–244.
- 690 Shipley, B. 2010. *From plant traits to vegetation structure. Chance and selection in the assembly*  
691 *of ecological communities*. Cambridge University Press, Cambridge, UK.
- 692 Šmilauer, P. & Lepš J. 2014. *Multivariate analysis of ecological data using CANOCO 5*. 2nd ed.  
693 Cambridge University Press, Cambridge, UK.
- 694 ter Braak, C.J.F, Peres-Neto, P. & Dray, S. 2017. A critical issue in model-based inference for  
695 studying trait-based community assembly and a solution. *PeerJ* 5:e2885.
- 696 ter Braak, C.J.F, Peres-Neto, P. & Dray, S. 2018. Simple parametric tests for trait-environment  
697 association. *Journal of Vegetation Science*, accepted.
- 698 ter Braak, C.J.F. & Barendregt, L.G. 1986. Weighted averaging of species indicator values: its  
699 efficiency in environmental calibration. *Mathematical Biosciences* 78: 57–72.

- 700 ter Braak, C.J.F. & Looman, C.W.N. 1986. Weighted averaging, logistic regression and the  
701 Gaussian response model. *Vegetatio* 65: 3–11.
- 702 ter Braak, C.J.F., Cormont, A. & Dray, S. 2012. Improved testing of species traits-environment  
703 relationships in the fourth-corner problem. *Ecology* 93: 1525–1526.
- 704 Vile, D., Shipley, B. & Garnier, E. 2006. Ecosystem productivity can be predicted from potential  
705 relative growth rate and species abundance. *Ecology Letters* 9: 1061–1067.
- 706 Westhoff, V., & van der Maarel, E. 1978. The Braun-Blanquet approach. In R. H. Whittaker  
707 (Ed.), *Classification of plant communities* (pp. 289–399). W. Junk, The Hague.
- 708 Wildi, O. 2016. Why mean indicator values are not biased. *Journal of Vegetation Science* 27:  
709 40–49.
- 710 Zelený, D. & Chytrý, M. 2007. Environmental control of vegetation pattern in deep river valleys  
711 of the Bohemian Massif. *Preslia*, 79: 205–222.
- 712 Zelený, D. & Schaffers, A.P. 2012. Too good to be true: pitfalls of using mean Ellenberg  
713 indicator values in vegetation analyses. *Journal of Vegetation Science* 23: 419–431.
- 714 Zelinka M. & Marvan P. 1961. Zur Präzisierung der biologischen Klassifikation der Reinheit  
715 fliessender Gewässer. *Archiv für Hydrobiologie* 57: 389–407.
- 716

717 **Tables with legends**

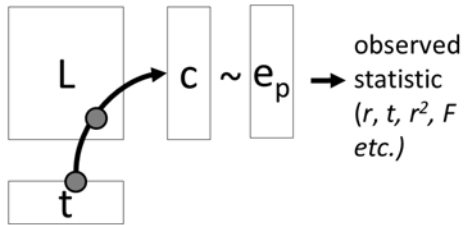
718 Table 1. Overview of the characteristics of the three categories of hypotheses tested by the CWM  
 719 approach. For each category, the corresponding assumption about a link between sample  
 720 attributes (**e**) or species attributes (**t**) and species composition (**L**) is provided (**X**  $\leftrightarrow$  **X**: no link,  
 721 **X**  $\leftrightarrow$  **X**: link), as well as the null vs alternative hypothesis and the recommended test.

Category of hypotheses	A	B	C
<b>The assumption about the link between objects (explicit or implicit)</b>	<b>t</b> $\leftrightarrow$ <b>L</b>	<b>e</b> $\leftrightarrow$ <b>L</b>	no assumption
<b>Null hypothesis</b>	<b>e</b> $\leftrightarrow$ <b>L</b>	<b>t</b> $\leftrightarrow$ <b>L</b>	<b>e</b> $\leftrightarrow$ <b>t</b> , i.e. <b>e</b> $\leftrightarrow$ <b>L</b> and/or <b>t</b> $\leftrightarrow$ <b>L</b>
<b>Alternative hypothesis</b>	<b>e</b> $\leftrightarrow$ <b>L</b>	<b>t</b> $\leftrightarrow$ <b>L</b>	<b>e</b> $\leftrightarrow$ <b>t</b> , i.e. <b>e</b> $\leftrightarrow$ <b>L</b> and <b>t</b> $\leftrightarrow$ <b>L</b>
<b>Recommended test</b>	<b>standard parametric test, row-based permutation test</b>	<b>a column-based permutation test (modified permutation test)</b>	<b>max test</b>

722

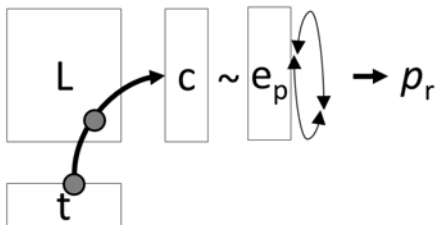
723 **Figures with legends**

**(a) Community-weighted mean approach**

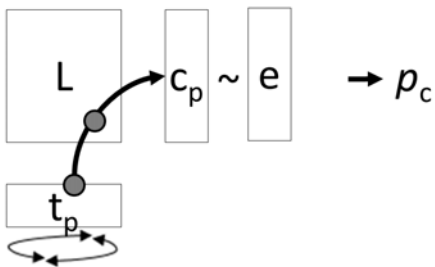


**(b) Tests of observed statistic**

(i) row-based permutation test ('standard test'):



(ii) column-based permutation test ('modified permutation test'):



(iii) row-column-based permutation test ('max test'):

$$p_{\max} = \max(p_r, p_c)$$

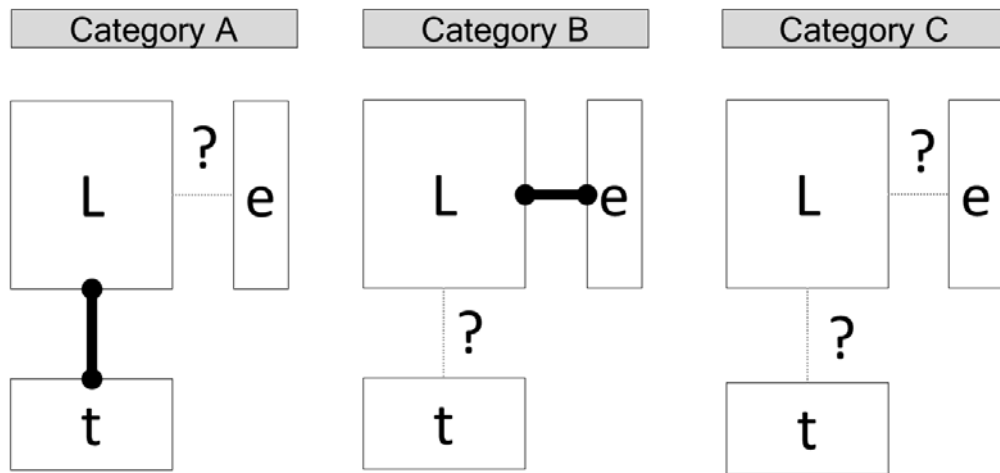
724

725 **Fig. 1.** Schema of (a) community-weighted mean approach, which generates the observed value

726 of the test statistic (depending on the method used), and (b) available tests of this statistic. Three

727 tests are available: (i) row-based permutation test (analogy to standard parametric test of c-e  
728 relationship), (ii) column-based permutation test (called ‘modified permutation test’ in Zelený &  
729 Schaffers 2012), and (iii) max test (called also ‘row-column based permutation test’ in Peres-  
730 Neto et al. 2017, or ‘sequential test’ in ter Braak et al. 2012). Notation:  $\mathbf{e}$  = sample attribute (e.g.  
731 environmental variable),  $\mathbf{t}$  = species attribute (e.g. trait),  $\mathbf{L}$  = matrix of species composition,  $\mathbf{c}$  =  
732 CWM calculated from  $\mathbf{t}$  and  $\mathbf{L}$ ,  $\mathbf{t}_p$  = species attributes permuted among species,  $\mathbf{e}_p$  = sample  
733 attributes permuted among samples,  $\mathbf{c}_p$  = CWM calculated from  $\mathbf{t}_p$  and  $\mathbf{L}$ .

734



735

736 **Fig. 2.** Differences in assumptions behind three categories of hypotheses tested by CWM

737 approach. The bold link indicates that the hypothesis explicitly or implicitly assumes the link

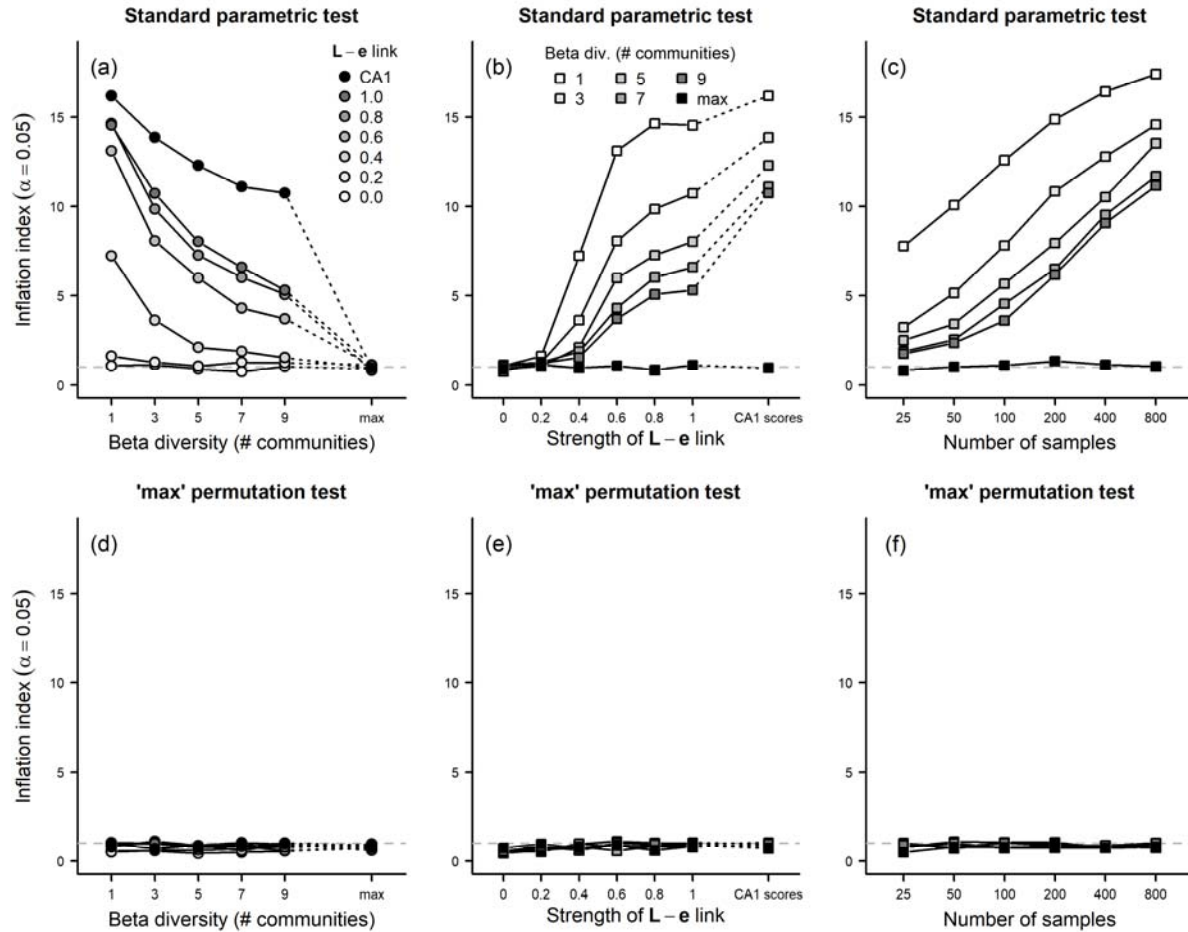
738 between the matrix of species composition (**L**) and either the vector of species attributes (**t**) or

739 vector sample attributes (**e**) and this link is therefore not tested. Question mark, on the other hand,

740 indicates that this link is not explicitly or implicitly acknowledged and can be tested.

741



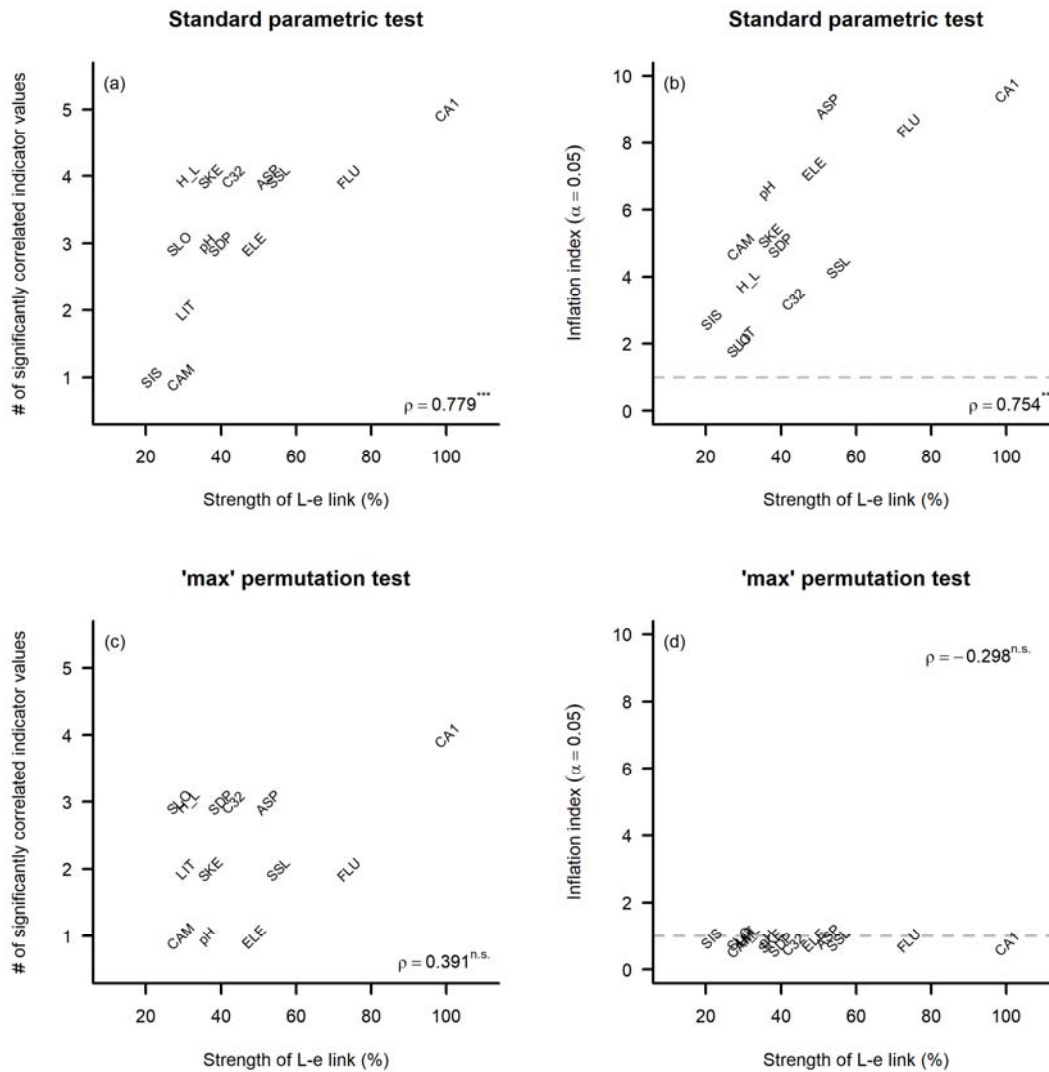


742

743 **Fig. 3.** The effect of data characteristics on the inflation index of CWM correlation between  
 744 CWM of species attributes and sample attributes, tested by parametric t-test of Pearson's  
 745 correlation coefficient (panels in top row) and max test (bottom row). Three data characteristics  
 746 were evaluated: beta diversity of the community dataset (number of communities 1-9 and max. =  
 747 maximum, when samples in the dataset does not share any species); the strength of the link  
 748 between sample attributes and species composition (**L-e**; 0 = no link, **e** completely randomized;  
 749 1 = full link, generated by the simulation model; CA1 = sample scores on the first CA axis), and  
 750 the number of samples in the community (25-800). CA1 scores are intrinsic sample attributes

751 with maximum strength of **L-e** link, since they are derived by correspondence analysis from the  
752 community matrix **L**. In each of the panels, one of the characteristics is fixed and the other two  
753 are left to vary: in (a, d) and (b, e) beta diversity and **L-e** link vary, while the number of samples  
754 is fixed ( $n = 100$ ), while in (c, f) the number of samples and beta diversity vary, while the  
755 strength of **L-e** is fixed (to value 0.6). The dashed horizontal line is for inflation index equal to  
756 one (no inflation).

757



758

759 Fig. 4. CWM correlations between CWM of Czech indicator values and environmental variables

760 in the real dataset (Vltava), tested by parametric t-test (a, b) and max permutation test (c, d). (a)

761 The number of significant ( $P < 0.05$ ) correlations between CWM of indicator values and

762 environmental variable (y-axis) increases with the strength of **L-e** relationship (x-axis, measured

763 as rescaled variance explained by this **e** in CCA). (b) Inflation index of CWM correlation (y-axis)

764 also increases with the strength of **L-e** relationship. If the max test replaces standard test, the

765 number of significant indicator values does not relate to the strength of **L-e** relationship (c) and

766 the inflation index is close to unity for all environmental variables (d).  $\rho$  = Spearman's  
767 coefficient of correlation between variables on  $x$ - and  $y$ -axis (\*\*\* =  $P < 0.001$ , \*\* =  $P < 0.01$ , n.s.  
768 = not significant). Dashed horizontal line (b, d) is for inflation index equal to one (no inflation).  
769 Environmental variables: ELE = elevation, SLO = slope, ASP = folded aspect, H\_L = heat load,  
770 SSL = landform shape in the downslope direction, SIS = landform shape along an isohypse, LIT  
771 = presence of lithic leptosols, SKE = presence of skeletal and hyperskeletal leptosols, CAM =  
772 presence of cambisol, FLU = presence of fluvisols, SDP = soil depth, pH = soil pH, C32 – cover  
773 of tree and shrub canopy.