

1 **rdacca.hp: an R package for generalizing hierarchical and variation**

2 **partitioning in multiple regression and canonical analysis**

3 Running title: **hierarchical partitioning in canonical analysis**

4 Jiangshan Lai^{1,2*}, Yi Zou³, Jinlong Zhang⁴, Pedro Peres-Neto⁵

5 ¹ *State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese*

6 *Academy of Sciences, Beijing, 100093, P.R. China*

7 ² *University of Chinese Academy of Sciences, Beijing, 100049, P.R. China*

8 ³ *Department of Health and Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou,*

9 *215123, P.R. China*

10 ⁴ *Flora Conservation Department, Kadoorie Farm and Botanic Garden, Lam Kam Road, Tai Po,*

11 *New Territories, Hong Kong SAR, China*

12 ⁵ *Department of Biology and Canada Research Chair in Biodiversity and Spatial Ecology,*

13 *Concordia University, Montreal, Quebec, Canada*

14 *Correspondence:

15 Jiangshan Lai (lai@ibcas.ac.cn)

16

17 **Summary**

- 18 1. Canonical analysis, a generalization of multiple regression to multiple response variables, is
19 widely used in ecology. Because these models often involve large amounts of parameters (one
20 slope per response per predictor), they pose challenges to model interpretation. Currently, multi-
21 response canonical analysis is constrained by two major challenges. Firstly, we lack quantitative
22 frameworks for estimating the overall importance of single predictors. Secondly, although the
23 commonly used variation partitioning framework to estimate the importance of groups of multiple
24 predictors can be used to estimate the importance of single predictors, it is currently
25 computationally constrained to a maximum of four predictor matrices.
- 26 2. We established that commonality analysis and hierarchical partitioning, widely used for both
27 estimating predictor importance and improving the interpretation of single-response regression
28 models, are related and complementary frameworks that can be expanded for the analysis of
29 multiple-response models.
- 30 3. In this application, we aim at: a) demonstrating the mathematical links between commonality
31 analysis, variation and hierarchical partitioning; b) generalizing these frameworks to allow the
32 analysis of any number of responses, predictor variables or groups of predictor variables in the
33 case of variation partitioning; and c) introducing and demonstrating the usage of the R package
34 `rdacca.hp` that implements these generalized frameworks.
- 35 Key-words: averaging over orderings, CCA, commonality analysis, constrained ordination,
36 explained variation, db-RDA, RDA, relative importance

37 **Introduction**

38 Canonical analysis (also called “constrained ordination”) considering multiple response
39 variables (e.g., species) and multiple predictors (e.g., environmental features, spatial predictors)
40 are widely used as inferential frameworks to determine and contrast the importance of multiple
41 drivers (e.g., environmental conditions, traits) underlying the structure of ecological communities.
42 Redundancy analysis (RDA; Rao 1964), canonical correspondence analysis (CCA; ter Braak
43 1986), and distance-based redundancy analysis (db-RDA; Legendre & Anderson 1999) are the
44 most commonly used ones (Legendre & Legendre 2012). One central challenge in canonical
45 analyses is the estimation of predictor contribution given that the number of regression parameters
46 increase as a function of the number of response variables. For instance, if 200 response variables
47 and 20 predictors are considered, one ends up with 4000 regression slopes.

48 Because canonical analyses are extensions of multiple regression models to multiple response
49 variables (Peres-Neto *et al.* 2006), we can adapt the existing machinery of multiple regression
50 models to tackle this challenge. The approach we develop here is based on generalizing
51 commonality analysis, a single-response regression framework, to canonical analysis (i.e., multi-
52 response variables). Commonality analysis is often used in psychology and education (Newton &
53 Spurell 1967; Nimon & Reio 2011; see Ray-Mukherjee *et al.* 2014 for a rare application in
54 ecology) to estimate the relative predictor contributions to the total model’s coefficient of
55 determination (R^2). Commonality analysis decomposes the total models’ R^2 into unique fractions
56 attributable to individual predictor and the shared fractions among predictors (covariation in the
57 predictive space which can also identify multicollinearity issues). These fractions are semi-partial
58 R^2 s (Peres-Neto *et al.* 2006) and allow circumventing well-known issues underlying variable

59 importance and model interpretation based solely on standardized partial slopes (see Ray-
60 Mukherjee *et al.* 2014 for a review of the issues). This is an interesting property because
61 canonical analyses produce estimates of total models' R^2 and semi-partial R^2 s that are unbiased
62 (Peres-Neto *et al.* 2006).

63 Although commonality analysis is applied to single-response regression models, ecologists
64 are widely familiar with the parallel framework of variation partitioning applied to canonical, with
65 thousands of studies published using it (Borcard, Legendre & Drapeau 1992; Peres-Neto *et al.*
66 2006; see Fig. 1). Variation partitioning is employed by grouping predictors together into matrices
67 and estimating the unique and shared semi-partial R^2 s of each matrix compounded across all
68 response variables (Peres-Neto *et al.* 2006). Indeed, the algebra involved in commonality analysis
69 and variation partitioning are equivalent even though they are often used in different situations.
70 Notwithstanding, packages conducting commonality analysis are restrained to single response
71 variables, e.g., R package `yhat` (Nimon, Oswald & Roberts 2013); and packages conducting
72 variation partitioning (i.e., multi-response), such as the widely used R package `vegan` (Oksanen
73 *et al.* 2006), are restrained to a maximum of four predictor matrices. These constraints are either
74 implementational or computational. In particular, variation partitioning has not been generalized to
75 multiple predictor matrices (but see Økland 2003) and commonality analysis has not been
76 generalized to multi-response models.

77 Beyond variation partitioning, estimating the relative importance of predictor in multiple
78 regression (e.g., multicollinearity being an extreme case) is an old and very active research (see
79 the reviews in Bi 2012; Nathans, Oswald & Nimon 2012; Grömping 2015). Among them, the
80 “averaging over orderings” approach proposed independently by Lindeman, Merenda & Gold

81 (1980), Cox (1985) and Kruskal (1987) in the 1980s has been considered as a breakthrough (Bi
82 2012). These methods are generally referred as to the LMG metric and are based on all model
83 subsets, and are also equal to methods described independently such as hierarchical partitioning
84 (Chevan & Sutherland 1991) and dominance analysis (Budescu 1993). Because the calculations
85 described in these papers are overly complicated, one is led to think that they differ. In the next
86 section, we simplify the presentation of “averaging over orderings” so that others can take full
87 advantage of its simplicity while describing model complexity. Our computational presentation
88 also makes a clear link between commonality analysis, and variation and hierarchical partitioning.
89 Because we were particularly inspired by the very popular paper by Chevan & Sutherland (1991),
90 we used the term “hierarchical partitioning” (HP) to relate to these equivalent methods. HP
91 produces all possible combinations of predictors to determine the order in which a predictor
92 dominates over the others (hence the name dominance analysis used by Badescu 1993) across all
93 subset models, and it has been widely used and recommended to assess the relative importance of
94 predictors in multiple regression (Soofi 1992; Mac Nally & Walsh 2004; Walsh *et al.* 2004;
95 Grömping 2006; 2007; 2009; 2015). Corresponding R packages are `relaimpo` for LMG
96 (Grömping 2006), `hier.part` for HP (Walsh & Mac Nally 2013) and `dominanceanalysis`
97 (Navarrete & Soares 2020). We show that analytical results from these packages are identical for
98 multiple regression (see Appendix S1). However, unlike variation partitioning, HP is not
99 currently available for multi-response models (i.e., canonical analysis) and is implemented in
100 our package: `rdacca.hp`. By doing so, we also extend HP to consider matrices of predictors
101 in variation partitioning.

102 The overall goal of this application paper is to unify commonality analysis, variation and

103 hierarchical partitioning, generalizing them to unlimited number of responses and predictor
104 variables (or matrices of predictors as in variation partitioning). These analyses are implemented
105 in our package `rdacca.hp` which is presented and illustrated in the next sections.

106 **Unifying commonality analysis, and variation and hierarchical partitioning**

107 Assuming a multi-response matrix \mathbf{Y} and three correlated predictors (Fig. 1). The fractions
108 of variation [a], [b] and [c] correspond to the variation in \mathbf{Y} that are uniquely explained by
109 predictors X_1 , X_2 , X_3 , respectively (i.e., unique semi-partial R^2). Fractions [d], [e], [f] are the
110 shared semi-partial R^2 s by combinations of their two respective predictors accounting for the third,
111 Fraction [g] is the shared semi-partial R^2 among all predictors, and [h] is the fraction
112 corresponding to residuals. All fractions summed totalize 100%. Shared fractions are the variation
113 in the response data that is explained by the correlation of the predictors involved. The larger this
114 fraction is, the more multicollinearity is present in the model. While model selection reduces the
115 shared variation among predictors (collinearity), it also reduces our ability to improve model
116 interpretability. The calculations involved in variation partitioning are well described elsewhere
117 (e.g., Peres-Neto *et al.* 2006) and our package: `rdacca.hp` generalizes it for any number of
118 predictors (or matrices of predictors).

119 Individual predictor contribution in HP (also called “independent contribution”) can be
120 simply estimated as its unique contribution to the total model R^2 plus its average shared
121 contributions with the other predictors, simplifying its presentation dramatically. The independent
122 contribution across all possible models derived from combinations of predictors can be derived
123 directly from variation partitioning (Fig. 1). For example, the independent contribution of X_1
124 (I_{X_1}) can be calculated by its unique contribution [a] and its respective shared contributions by

125 the number of predictors involved in each of:

$$126 \quad I_{X_1} = a + \frac{d}{2} + \frac{f}{2} + \frac{g}{3} \quad (1)$$

127 For completion, the contributions of X_2 and X_3 are calculated as follows:

$$128 \quad I_{X_2} = b + \frac{d}{2} + \frac{e}{2} + \frac{g}{3} \quad (2)$$

$$129 \quad I_{X_3} = c + \frac{f}{2} + \frac{e}{2} + \frac{g}{3} \quad (3)$$

130 and:

$$131 \quad I_{X_1} + I_{X_2} + I_{X_3} = R_{Total}^2 \quad (4)$$

132 Generalizing, the independent contribution of any predictor i (I_{X_i}) can be computed as:

$$133 \quad I_{X_i} = \sum_{k=1}^n \sum_{j=1}^m \frac{R_{SX_{ik,j}}^2}{k} \quad (5)$$

134 where n is the number of predictors, $R_{SX_{ik,j}}^2$ is semi-partial R^2 of the j^{th} fraction shared between X_i

135 and the other k predictors, and m is the number of combinations that X_i shared with other k

136 predictors, with $m = \binom{k-1}{n-1}$. Note that as the number of predictor increases, the number of

137 fractions including shared and unique R^2 s increases exponentially (2^N-1 fractions). The above

138 computation only uses individual predictors to illustrate the process. It is also applicable to

139 matrices of predictors as in routine ecological applications applying variation partitioning.

140 In general, we find variation partitioning should be the starting point prior to hierarchical

141 partitioning. While the former emphasizes unique and common variation among predictors, the

142 latter emphasizes the overall importance of each predictor (or group of predictors). Our package

143 `rdacca.hp` synchronously implements variation and hierarchical partitioning for single- and

144 multiple-response models without limits in the number of predictors / matrices of predictors

145 **Package description**

146 The `rdacca.hp` package is written in R (R Development Core Team 2019) and can be

147 installed from CRAN (<http://cran.r-project.org/web/packages/rdacca.hp/>) or Github
148 (<https://github.com/laijiangshan/rdacca.hp>). The package contains one key homonymous function:
149 `rdacca.hp` that conducts both variation and hierarchical partitioning in single- and multiple-
150 response multiple regression (canonical analysis). The internal function: `Canonical.Rsq`,
151 calculates R^2 and adjusted R^2 (hereafter R^2_{adj}) of RDA, db-RDA and CCA, which are called by
152 `rdacca.hp`. For canonical analysis, the R^2_{adj} is used given that the contribution of null
153 predictors can differ quite a lot from zero due to sampling variation related to large number of
154 predictors and small number of observations (Peres-Neto *et al.* 2006). The R^2_{adj} is calculated
155 using Ezekiel's formula (Ezekiel 1930) for RDA and db-RDA, while permutation procedure be
156 used for CCA (Peres-Neto *et al.* 2006). The interpretation of arguments and some key notes in
157 function `rdacca.hp` are described briefly here.
158 `rdacca.hp(dv, iv, method, type, n.perm, trace, plot.perc)`
159 In this usage, both a variation partitioning and hierarchical partitioning are performed in which the
160 unique, shared (referred as to "common") and independent contributions of each predictor
161 (columns in `iv`) to the global R^2 canonical model are computed. In the section below (working
162 examples), we show how the function can conduct the analysis based on blocks of variables as in
163 routine ecological applications of variation partitioning. Response variables (columns in `dv`) must
164 be numerical, while the predictors (`iv`) can be either numerical or categorical. `method` is the type
165 of canonical analysis set as either RDA (default), dbRDA, or CCA. If `method="dbRDA"`, `dv`
166 should be a distance matrix (i.e., `dist` class). If `dv` is imputed as one numerical vector, RDA
167 would be equivalent to the classic (single response) multiple regression (see Appendix S1). An
168 additional advantage of our package in relation to `relaimpo`, `hier.part` and

169 `dominanceanalysis` is that it also decomposes R^2_{adj} for multiple regression, so this package
170 is also useful across many other ecological applications and research fields. `type` is the type of
171 total explained variation: "`adjR2`" is for R^2_{adj} and "`R2`" for unadjusted R^2 , by default is
172 "`adjR2`". `n.perm` is the number of permutations when computing R^2_{adj} for CCA and default is
173 1000 to get a relatively stable value. `trace` is a logical argument indicating whether the output of
174 variation partitioning should be printed (see the example below). It is set by default as `FALSE` to
175 save screen space, as this output will increase exponentially with the number of predictors. If
176 focusing on variation partitioning, then `trace=TRUE` should be set. `plot.perc` is a logical
177 argument indicating whether a bar plot of the relative independent contribution of predictors is
178 plotted; the default is `FALSE`. The output of `rdacca.hp` is explained in the example below.

179 **A working example**

180 We illustrate the usage of `rdacca.hp` by using the Doubs River fish data readily available
181 in the `ade4` package (Thioulouse *et al.* 2018). The dataset is a subset of the data originally
182 collected by Verneaux (1973) with the distributions of fish species and environmental factors
183 along the Doubs River in the Jura Mountains, near the France–Switzerland border (also see
184 Verneaux *et al.* 2003). This dataset contains 27 fish species with abundance classes (ranging from
185 0 to 5) and 11 quantitative environmental variables describing the river morphology and water
186 quality from 30 sites.

187 We show the results for the variation and hierarchical partitioning of the global R^2_{adj} of
188 RDA for the Doubs fish data via `rdacca.hp` for three predictors: *alt* (altitude), *oxy* (dissolved
189 oxygen) and *bdo* (biological demand for oxygen) to simplify the presentation of the package rather
190 than provide a complete ecological analysis and related interpretations. These variables (again for

191 presentation convenience) were selected using a stepwise selection based on all environmental
192 predictors using the function `ordistep` in the `vegan` package (Oksanen *et.al.* 2019). Note that
193 the selection of predictors is not a prerequisite of `rdacca.hp` and, as previously noted earlier,
194 can generate biases and incomplete information in variation and hierarchical partitioning. Our
195 function `rdacca.hp` does not limit the number of predictors or groups. Note though, that
196 considering all combinations of all variables is important even though computationally
197 demanding. The selected predictors have a relatively strong correlation structure: a relatively high
198 negative correlation ($r=-0.84$, $p<0.001$) between *oxy* and *bdo*; a relatively weaker positive
199 correlation between *alt* and *oxy* ($r=0.42$, $p=0.02$), and a negative correlation between *alt* and *bdo*
200 ($r=-0.38$, $p=0.04$). This is an important point to make (often missed by practitioners) because the
201 correlation structure in the predictor space (simple pairwise correlations between predictors) may
202 not translate necessarily into high shared fractions among predictors (i.e., semi-partial R^2
203 calculated on the basis of the predictive space). As such, contrasting the correlation structure
204 among predictors and their correlation structure in predictive space (i.e., variation and hierarchical
205 partitioning) should serve useful in generating insights underlying the relational nature of
206 predictors and responses.

207 The script below demonstrates the use and outputs of `rdacca.hp` to decompose the
208 global R^2_{adj} in RDA. Decomposition of R^2_{adj} in CCA and db-RDA are similar to RDA and
209 shown in the Appendix S2. The standard output of `rdacca.hp` is as follows:

```
210 #install rdacca.hp from CRAN  
211 install.packages('rdacca.hp')  
212 library(rdacca.hp)  
  
213 #sample data in ade4 package  
214 require(ade4)
```

```
215 data(doubs)
216 #fish species as response variables
217 spe <- doubs$fish
218 #environmental factors as explanatory variables
219 env <- doubs$env
220 #remove empty site 8 without species
221 spe <- spe[-8,]
222 env <- env[-8,]
223 #'dfs' is a variable containing locations rather than environmental
224 #factor. We also removed it from the 'env' data frame
225 env <- env[, -1]
226 #the usual Hellinger-transformation of the species dataset
227 spe.hel <- decostand(spe, "hellinger")
228 #selecting variables via ordistep() in vegan
229 ordistep(rda(spe.hel~., env))

230 #three selected variables: alt, oxy and bdo
231 rdacca.hp(spe.hel, env[, c("alt", "oxy", "bdo")], method="RDA", type = "a
232 djR2", trace = TRUE, plot.perc = FALSE)

233 ## $Method_Type
234 ## [1] "RDA" "adjR2"
235 ##
236 ## $R.squared
237 ## [1] 0.5402
238 ##
239 ## $Var.part
240 ##           Fractions  % Total
241 ## Unique to alt           0.1942   35.96
242 ## Unique to oxy           0.1367   25.31
243 ## Unique to bdo           0.0871   16.13
244 ## Common to alt, and oxy   0.0467    8.64
245 ## Common to alt, and bdo  -0.0179   -3.31
246 ## Common to oxy, and bdo   0.0131    2.43
247 ## Common to alt, oxy, and bdo 0.0801   14.84
248 ## Total                   0.5402  100.00
249 ##
250 ## $Hier.part
251 ##   Independent I.perc(%)
252 ## alt    0.2354   43.58
253 ## oxy    0.1933   35.78
254 ## bdo    0.1115   20.64
```

255 The function `rdacca.hp` returns a list containing three bits (if `trace=FALSE`) or four bits (if

256 `trace=TRUE`).

257 `$Method_Type`: This bit shows the type of canonical analysis and whether the original or adjusted

258 R^2 were used in the analysis. In the example, it is RDA and `adjR2`.

259 `$R.squared`: `type="adjR2"`. In this example, the amount of variation of the fish data matrix

260 explained by the three predictors (**alt**, **oxy** and **bdo**) is 54.02%.

261 `$Var.part`: It contains output that lists with R^2 adj values for all fractions based on variation

262 partitioning (commonality analysis) if setting `trace=TRUE`. In this example, with three

263 predictors, commonality analysis decomposes R^2 adj into all seven fractions (i.e., 2^3-1) of unique

264 and common effects. Fractions can be either positive or negative (Nimon & Reio 2011, Ray-

265 Mukherjee *et al* 2014). In this case, the common effect between *alt* and *bod* is negative (-0.0179).

266 Negative common (shared) variation is possible when predictors act as suppressors of other

267 predictors (Pedhazur 1997; Nimon & Reio 2011; Ray-Mukherjee *et al* 2014).

268 All fractions (except the residual) sum to the total R^2 adj (i.e., the value of `$R.squared`).

269 `$Hier.part`: This bit is a matrix containing the independent contribution of each predictor

270 in the “Independent” column, and their percentage in the “I.perc” column. In this case, the

271 independent contributions of *alt*, *oxy* and *bdo* are 0.2354, 0.1933 and 0.1115, respectively; note

272 again that they sum to the overall R^2 adj (0.5402). One easily know how to calculate the

273 independent contribution based the result of variation partitioning. For instance, the value

274 (0.1933) for *oxy* is the sum of its unique effect and three average common effects (i.e., $0.1367+$

275 $0.0467/2 + 0.0131/2 + 0.0801/3$). `rdacca.hp` also produces a bar graph based on the `ggplot2`

276 package (Wickham 2016) (Fig. 2). Users are encouraged to use other graph functions to generate

277 other types of plots based on the numerical output of `rdacca.hp`.

278 The following code shows how to apply `rdacca.hp` to conduct a variation partitioning and
279 hierarchical partitioning to two sets of matrices containing environmental variables describing
280 river morphology and water quality. Note `rdacca.hp` has no limit for the number of sets of
281 predictor matrices and an example with five predictor groups is provided in the Appendix S3.
282 `rdacca.hp` does not include the traditional representation of variation partitioning as a Venn
283 diagram, which becomes impossible for more than four predictors.

```
284 rdacca.hp(spe.hel, list(envtopo=env[,1:3], envchem=env[,4:10]), method  
285 ="RDA", type = "adjR2", trace = TRUE, plot.perc = TRUE)  
286 ## $Method_Type  
287 ## [1] "RDA" "adjR2"  
288 ##  
289 ## $R.squared  
290 ## [1] 0.5558  
291 ##  
292 ## $Var.part  
293 ##  
294 ## Unique to envtopo          Fractions  % Total  
295 ## Unique to envchem          0.2290    41.21  
296 ## Common to envtopo, and envchem 0.2454    44.15  
297 ## Total                      0.5558   100.00  
298 ##  
299 ## $Hier.part  
300 ##           Independent I.perc(%)  
301 ## envtopo      0.2041    36.72  
302 ## envchem      0.3517    63.28
```

303 Discussion

304 Multivariate regressions and canonical analysis are quintessential quantitative frameworks to
305 tackle many ecological and evolutionary problems. Variation partitioning and hierarchical
306 partitioning (HP) are frameworks that allow going beyond the usual and standard interpretation of
307 partial slopes (see Ray-Mukherjee *et al.* 2014 for a review). That said, our presentation of HP
308 allows realizing that both methods are interrelated, and that variation partitioning can serve as a

309 precursor of HP. While variation partitioning estimates unique and common variation in a single
310 full model containing all variables, HP is based on principles of all subset regression and model
311 averaging which are known to improve model inference and interpretability over traditional model
312 selection procedures (Burnham & Anderson 2002). The two main motivations for using variation
313 partitioning and HP over traditional model selection procedures are that: a) candidate predictors
314 are conditional on the variance explained by the models already retained by the selection
315 procedure (Thompson 1995; Nathans, Oswald & Nimon 2012). As such, a predictor that can be
316 quite relevant in multiple sub-models (indicating overall importance) may not be retained by
317 model selection procedures; b) model selection can be heavily influenced by sampling variation
318 (Thompson 1995; Nathans, Oswald & Nimon 2012); if another sample were to be used, the
319 selected predictors could vary. By using HP, one can analyze variable importance over all possible
320 models.

321 While most canonical analysis is used to analyze species distributional matrices, our package
322 can also motivate ecologists to explore its use to different types of response matrices and problems
323 (e.g., multiple traits). Note that we decided not to consider significance testing for independent
324 contributions in HP because predictor (or groups of predictors) relative importance is usually
325 considered as an exploratory framework for interpreting regression rather than an inferential tool.
326 Finally, one disadvantage of the variation partitioning and HP frameworks is that the calculation
327 volume increases exponentially with the increase of predictor variables. We will continue to
328 optimize our package to improve, among other issues, the calculation speed. Although we
329 demonstrated the application of `rdacca.hp` package using an ecology example, this package is
330 also applicable to any other areas where canonical analysis and multiple regressions are applied.

331 Currently, our `rdacca.hp` package has been used in peer-reviewed papers (e.g. Li *et al.* 2020;

332 Song *et al.* 2020; Sun *et al.* 2020; Wang *et al.* 2020; Xiong *et al.* 2020; Zhou *et al.* 2020).

333 Researchers using the `rdacca.hp` package in their studies, should cite this article and the

334 `rdacca.hp` package as well. Citation information can be obtained by typing:

335 `citation("rdacca.hp")`.

336 **Acknowledgements**

337 The research was supported by the Strategic Priority Research Program of the Chinese Academy

338 of Sciences (XDA19050404) and the National Science and Technology Basic Resources Survey

339 Program of China (2019FY100204). PP-N was supported by the Canada Research Chair (CRC)

340 program. Otherwise there is no conflict of interest among the authors.

341 **Authors' contributions**

342 JSL conceived the idea. JSL and PP-N wrote the package and conducted the analysis. All authors

343 participated in writing multiples drafts.

344 **References**

345 Bi, J. (2012) A review of statistical methods for determination of relative importance of correlated

346 predictors and Identification of drivers of consumer liking. *Journal of Sensory Studies*, **27**, 87-

347 101.

348 Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological

349 variation. *Ecology*, **73**, 1045-1055.

350 Budescu, D.V. (1993) Dominance analysis - a new approach to the problem of relative importance of

351 predictors in multiple-regression. *Psychological Bulletin*, **114**, 542-551.

- 352 Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference : a practical*
353 *information-theoretic approach*, second edn. Springer-Verlag, New York.
- 354 Chevan, A. & Sutherland, M. (1991) Hierarchical partitioning. *American Statistician*, **45**, 90-96.
- 355 Cox, L.A. (1985) A new measure of attributable risk for public health applications. *Management*
356 *Science*, **31**, 800-813.
- 357 Ezekiel, M. (1930) *Methods of Correlational Analysis*. Wiley, New York
- 358 Grömping, U. (2006) Relative importance for linear regression in R: The package relaimpo. *Journal of*
359 *Statistical Software*, **17**, 1-27..
- 360 Grömping, U. (2007) Relative importance in linear regression based on variance decomposition.
361 *American Statistician*, **61**, 139-147.
- 362 Grömping, U. (2009) Variable importance assessment in regression: Linear regression versus random
363 forest. *American Statistician*, **63**, 308-319.
- 364 Grömping, U. (2015) Variable importance in regression models. *WIREs Computational Statistics*, **7**,
365 137-152.
- 366 Kruskal, W. (1987) Relative importance by averaging over orderings. *American Statistician*, **41**, 6-10.
- 367 Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: Testing multispecies
368 responses in multifactorial ecological experiments. *Ecological Monographs*, **69**, 1-24.
- 369 Legendre, P. & Legendre, L. (2012) *Numerical Ecology*, Third edn. Elsevier.
- 370 Li, J.Y., Wang, N.L., Dodson, J., Yan, H., Zhang, X.J., Jia, P.W. & Seppa, H. (2020) Holocene negative
371 coupling of summer temperature and moisture availability over southeastern arid Central Asia.
372 *Climate Dynamics*, **55**, 1187-1208.
- 373 Lindeman, R.H., Merenda, P.F. & Gold, R.Z. (1980) *Introduction to Bivariate and Multivariate*

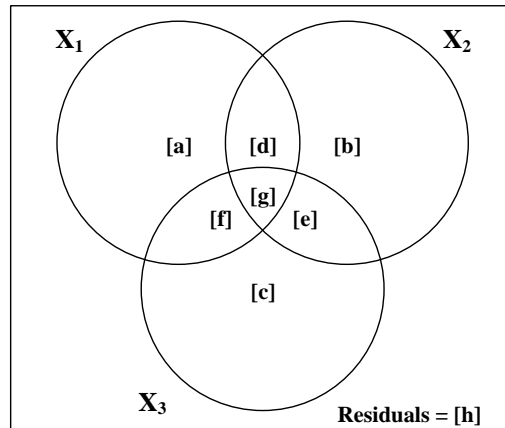
- 374 *Analysis*. Scott Foresman, Glenview, IL.
- 375 Mac Nally, R. & Walsh, C.J. (2004) Hierarchical partitioning public-domain software. *Biodiversity and*
376 *Conservation*, **13**, 659-660.
- 377 Nathans, L.L., Oswald, F.L. & Nimon, K. (2012) Multiple linear regression: A guidebook of variable
378 importance. *Practical Assessment, Research & Evaluation*, **17**, 1-19.
- 379 Navarrete, C.B. & Soares, F.C. (2020) dominanceanalysis: Dominance Analysis. R package version
380 2.0.0.
- 381 Newton, R.G. & Spurrell, D.J. (1967) Examples of the use of elements for classifying regression
382 analysis. *Applied Statistics*, **16**, 165-172.
- 383 Nimon, K. & Reio, T. (2011) Regression commonality analysis: a technique for quantitative theory
384 building. *Human Resource Development*, **10**, 329– 340.
- 385 Nimon, K., Oswald, F.L. & Roberts, J.K. (2013) Yhat: Interpreting regression effects. R package
386 version 2.0.0.
- 387 Økland, R.H. (2003) Partitioning the variation in a plot-by-species data matrix that is related to n sets
388 of explanatory variables. *Journal of Vegetation Science*, **14**, 693-700.
- 389 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara,
390 R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2019) vegan:
391 Community Ecology Package. R package version 2.5-4.
- 392 Oksanen, J., Kindt, R., Legendre, P. & O'Hara, R.B. (2006) vegan: Community Ecology Package. R
393 package version 1.8-1.
- 394 Pedhazur, E.J. (1997) Multiple Regression in Behavioral Research: Explanation and Prediction, 3rd
395 edn. Harcourt Brace, Orlando, FL.

- 396 Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data
397 matrices: Estimation and comparison of fractions. *Ecology*, **87**, 2614-2625.
- 398 R Development Core Team (2019) R: A Language and Environment for Statistical Computing. R
399 Foundation for Statistical Computing, Vienna, Austria.
- 400 Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research.
401 *Sankhyā A*, **26**, 329-358.
- 402 Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. (2014) Using
403 commonality analysis in multiple regressions: a tool to decompose regression effects in the
404 face of multicollinearity. *Methods in Ecology and Evolution*, **5**, 320-328.
- 405 Song, S.S., Zhang, C., Gao, Y., Zhu, X.Y., Wang, R.H., Wang, M.D., Zheng, Y.L., Hou, L.J., Liu, M. &
406 Wu, D.M. (2020) Responses of wetland soil bacterial community and edaphic factors to two-
407 year experimental warming and *Spartina alterniflora* invasion in Chongming Island. *Journal*
408 *of Cleaner Production*, **250**, doi.org/10.1016/j.jclepro.2019.119502.
- 409 Soofi, E.S. (1992) A generalizable formulation of conditional logit with diagnostics. *Journal of the*
410 *American Statistical Association*, **87**, 812-816.
- 411 Sun, X.J., Diekman, M., Yan, X., Zou, Y., Sang, W.G. & Axmacher, J.C. (2020) Diversity and seasonal
412 changes in carabid assemblages of a mature, secondary and plantation forest mosaic in the
413 Zhangguangcai Mountains in northeastern China. *Insect Conservation and Diversity*, **13**, 340-
414 350.
- 415 ter Braak, C.J.F. (1986) Canonical correspondence-analysis - a new eigenvector technique for
416 multivariate direct gradient analysis. *Ecology*, **67**, 1167-1179.
- 417 Thioulouse, J., Dray, S., Dufour, A.B., Siberchicot, A., Jombart, T. & Pavoine, S. (2018) *Multivariate*

- 418 *Analysis of Ecological Data with ade4*. Springer.
- 419 Thompson, B. (1995) Stepwise regression and stepwise discriminant analysis need not apply here: A
420 guidelines editorial. *Educational and Psychological Measurement*, **55**, 525-534.
- 421 Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur*
422 *le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse d'état, Besançon.
- 423 Verneaux, J., Schmitt, A., Verneaux, V. & Prouteau, C. (2003) Benthic insects and fish of the Doubs
424 River system: typological traits and the development of a species continuum in a theoretically
425 extrapolated watercourse. *Hydrobiologia*, **490**, 63-74.
- 426 Walsh, C.J. & Mac Nally, R. (2013) hier.part: Hierarchical Partitioning. R package version 1.0-4.
- 427 Walsh, C.J., Papas, P.J., Crowther, D. & Yoo, J. (2004) Stormwater drainage pipes as a threat to a
428 stream-dwelling amphipod of conservation significance, *Austrogammarus australis*, in
429 southeastern Australia. *Biodiversity and Conservation*, **13**, 781-793.
- 430 Wang, Y., Yang, X.D., Ali, A., Lv, G.H., Long, Y.X., Wang, Y.Y., Ma, Y.G. & Xu, C.C. (2020)
431 Flowering phenology shifts in response to functional traits, growth form, and phylogeny of
432 woody species in a desert area. *Frontiers in Plant Science*, **11**, doi: 10.3389/fpls.2020.00536
- 433 Wickham, H. (2016) *Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- 434 Xiong, Q.L., Luo, X.J., Liang, P.H., Xiao, Y., Xiao, Q., Sun, H., Pan, K.W., Wang, L.X., Li, L.J. &
435 Pang, X.Y. (2020) Fire from policy, human interventions, or biophysical factors? Temporal-
436 spatial patterns of forest fire in southwestern China. *Forest Ecology and Management*, **474**,
437 doi.org/10.1016/j.foreco.2020.118381
- 438 Zhou, S.L., Sun, Y., Li, Z.X. & Huang, T.L. (2020) Characteristics and driving factors of the aerobic
439 denitrifying microbial community in Baiyangdian Lake, Xiong'an New Area. *Microorganisms*,

440 **8**, doi:10.3390/microorganisms8050714

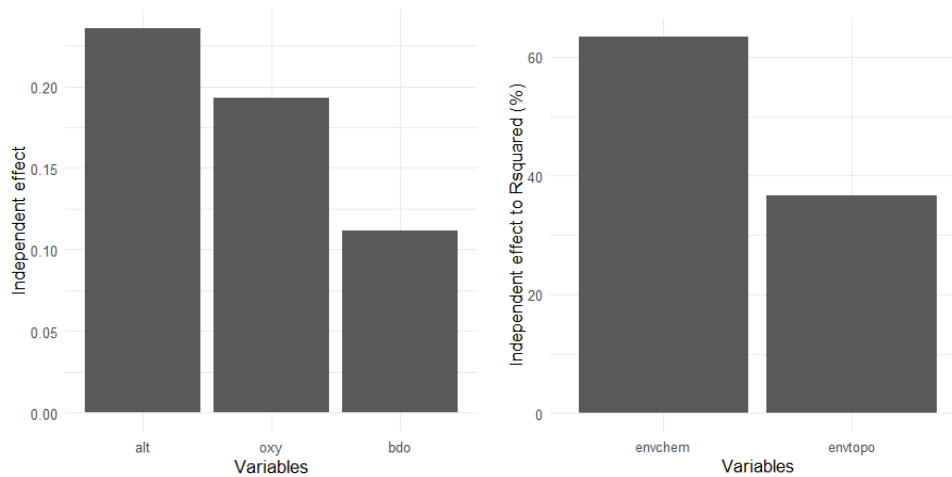
441



442

443 Figure 1. Venn diagram representing the variation partitioning of a response matrix Y regressed
444 against three correlated predictors (or groups of predictors as in variation partitioning). All
445 fractions add to 100% and the bounding rectangle represents the total variation in the response
446 data (i.e., 100%) while each circle represents the relative portion of variation accounted by
447 different fractions (see text for a detailed calculations and further explanation).

448



449

450 Figure 2. The graphic output of `rdacca.hp` function for the Doubs River fish data. left panel:
451 original independent contribution for three predictors (`plot.perc=FLASE`) and right panel:
452 their relative contribution for two groups to the overall adjusted R^2 (`plot.perc=TRUE`).