`rdacca.hp`: an R package for generalizing hierarchical and variation partitioning in multiple regression and canonical analysis

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

Jiangshan Lai[1,2*], Yi Zou[3], Jinlong Zhang[4], Pedro Peres-Neto[5]

[1] *State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, P.R. China*

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

[3] *Department of Health and Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, P.R. China*

[4] *Flora Conservation Department, Kadoorie Farm and Botanic Garden, Lam Kam Road, Tai Po, New Territories, Hong Kong SAR, China*

[5] *Department of Biology and Canada Research Chair in Biodiversity and Spatial Ecology, Concordia University, Montreal, Quebec, Canada*

[*]Correspondence:

Jiangshan Lai (lai@ibcas.ac.cn)

Summary

1. Canonical analysis, a generalization of multiple regression to multiple response variables, is widely used in ecology. Because these models often involve large amounts of parameters (one slope per response per predictor), they pose challenges to model interpretation. Currently, multi-response canonical analysis is constrained by two major challenges. Firstly, we lack quantitative frameworks for estimating the overall importance of single predictors. Secondly, although the commonly used variation partitioning framework to estimate the importance of groups of multiple predictors can be used to estimate the importance of single predictors, it is currently computationally constrained to a maximum of four predictor matrices.

2. We established that commonality analysis and hierarchical partitioning, widely used for both estimating predictor importance and improving the interpretation of single-response regression models, are related and complementary frameworks that can be expanded for the analysis of multiple-response models.

3. In this application, we aim at: a) demonstrating the mathematical links between commonality analysis, variation and hierarchical partitioning; b) generalizing these frameworks to allow the analysis of any number of responses, predictor variables or groups of predictor variables in the case of variation partitioning; and c) introducing and demonstrating the usage of the R package `rdacca.hp` that implements these generalized frameworks.

Key-words: averaging over orderings, CCA, commonality analysis, constrained ordination, explained variation, db-RDA, RDA, relative importance

Introduction

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

Because canonical analyses are extensions of multiple regression models to multiple response variables (Peres-Neto *et al.* 2006), we can adapt the existing machinery of multiple regression models to tackle this challenge. The approach we develop here is based on generalizing commonality analysis, a single-response regression framework, to canonical analysis (i.e., multi-response variables). Commonality analysis is often used in psychology and education (Newton & Spurell 1967; Nimon & Reio 2011; see Ray-Mukherjee *et al.* 2014 for a rare application in ecology) to estimate the relative predictor contributions to the total model's coefficient of determination (R^2). Commonality analysis decomposes the total models' R^2 into unique fractions attributable to individual predictor and the shared fractions among predictors (covariation in the predictive space which can also identify multicollinearity issues). These fractions are semi-partial R^2s (Peres-Neto *et al.* 2006) and allow circumventing well-known issues underlying variable importance and model interpretation based solely on standardized partial slopes (see Ray-Mukherjee *et al.* 2014 for a review of the issues). This is an interesting property because canonical analyses produce estimates of total models' R^2 and semi-partial R^2s that are unbiased (Peres-Neto *et al.* 2006).

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

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

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

The overall goal of this application paper is to unify commonality analysis, variation and hierarchical partitioning, generalizing them to unlimited number of responses and predictor variables (or matrices of predictors as in variation partitioning). These analyses are implemented in our package `rdacca.hp` which is presented and illustrated in the next sections.

Unifying commonality analysis, and variation and hierarchical partitioning

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

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

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

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

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

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

and:

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

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

$$I_{X_i} = \sum\nolimits_{k=1}^{n} \sum\nolimits_{j=1}^{m} \frac{R^2_{SX_{i,k,j}}}{k} \quad (5)$$

where n is the number of predictors, $R^2_{SX_{i,k,j}}$ is semi-partial R^2 of the j^{th} fraction shared between X_i and the other k predictors, and m is the number of combinations that X_i shared with other k predictors, with $m = \binom{k-1}{n-1}$. Note that as the number of predictor increases, the number of fractions including shared and unique R^2s increases exponentially (2^N-1 fractions). The above computation only uses individual predictors to illustrate the process. It is also applicable to matrices of predictors as in routine ecological applications applying variation partitioning.

In general, we find variation partitioning should be the starting point prior to hierarchical partitioning. While the former emphasizes unique and common variation among predictors, the latter emphasizes the overall importance of each predictor (or group of predictors). Our package `rdacca.hp` synchronously implements variation and hierarchical partitioning for single- and multiple-response models without limits in the number of predictors / matrices of predictors

Package description

The `rdacca.hp` package is written in R (R Development Core Team 2019) and can be installed from CRAN (http://cran.r-project.org/web/packages/rdacca.hp/) or Github (https://github.com/laijiangshan/rdacca.hp). The package contains one key homonymous function: `rdacca.hp` that conducts both variation and hierarchical partitioning in single- and multiple-response multiple regression (canonical analysis). The internal function: `Canonical.Rsq`, calculates R^2 and adjusted R^2 (hereafter R^2adj) of RDA, db-RDA and CCA, which are called by `rdacca.hp`. For canonical analysis, the R^2adj is used given that the contribution of null predictors can differ quite a lot from zero due to sampling variation related to large number of predictors and small number of observations (Peres-Neto *et al.* 2006). The R^2adj is calculated using Ezekiel's formula (Ezekiel 1930) for RDA and db-RDA, while permutation procedure be used for CCA (Peres-Neto *et al.* 2006). The interpretation of arguments and some key notes in function `rdacca.hp` are described briefly here.

```
rdacca.hp(dv, iv, method, type, n.perm, trace, plot.perc)
```

In this usage, both a variation partitioning and hierarchical partitioning are performed in which the unique, shared (referred as to "common") and independent contributions of each predictor (columns in `iv`) to the global R^2 canonical model are computed. In the section below (working examples), we show how the function can conduct the analysis based on blocks of variables as in routine ecological applications of variation partitioning. Response variables (columns in `dv`) must be numerical, while the predictors (`iv`) can be either numerical or categorical. `method` is the type of canonical analysis set as either `RDA` (default), `dbRDA`, or `CCA`. If `method=`"dbRDA", `dv` should be a distance matrix (i.e., `dist` class). If `dv` is imputed as one numerical vector, RDA would be equivalent to the classic (single response) multiple regression (see Appendix S1). An additional advantage of our package in relation to `relaimpo`, `hier.part` and `dominanceanalysis` is that it also decomposes R^2adj for multiple regression, so this package is also useful across many other ecological applications and research fields. `type` is the type of total explained variation: "`adjR2`" is for R^2adj and "`R2`" for unadjusted R^2, by default is "`adjR2`". `n.perm` is the number of permutations when computing R^2adj for CCA and default is 1000 to get a relatively stable value. `trace` is a logical argument indicating whether the output of variation partitioning should be printed (see the example below). It is set by default as FALSE to save screen space, as this output will increase exponentially with the number of predictors. If focusing on variation partitioning, then `trace=TRUE` should be set. `plot.perc` is a logical argument indicating whether a bar plot of the relative independent contribution of predictors is plotted; the default is `FALSE`. The output of `rdacca.hp` is explained in the example below.

A working example

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

We show the results for the variation and hierarchical partitioning of the global R^2adj of RDA for the Doubs fish data via `rdacca.hp` for three predictors: *alt* (altitude), *oxy* (dissolved oxygen) and *bdo* (biological demand for oxygen) to simplify the presentation of the package rather than provide a complete ecological analysis and related interpretations. These variables (again for presentation convenience) were selected using a stepwise selection based on all environmental predictors using the function `ordistep` in the `vegan` package (Oksanen *et.al.* 2019). Note that the selection of predictors is not a prerequisite of `rdacca.hp` and, as previously noted earlier, can generate biases and incomplete information in variation and hierarchical partitioning. Our function `rdacca.hp` does not limit the number of predictors or groups. Note though, that considering all combinations of all variables is important even though computationally demanding. The selected predictors have a relatively strong correlation structure: a relatively high negative correlation ($r=-0.84$, $p<0.001$) between *oxy* and *bdo*; a relatively weaker positive correlation between *alt* and *oxy* ($r=0.42$, $p=0.02$), and a negative correlation between *alt* and *bdo* ($r=-0.38$, $p=0.04$). This is an important point to make (often missed by practitioners) because the correlation structure in the predictor space (simple pairwise correlations between predictors) may not translate necessarily into high shared fractions among predictors (i.e., semi-partial R^2 calculated on the basis of the predictive space). As such, contrasting the correlation structure among predictors and their correlation structure in predictive space (i.e., variation and hierarchical partitioning) should serve useful in generating insights underlying the relational nature of predictors and responses.

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

```
#install rdacca.hp from CRAN
install.packages('rdacca.hp')
library(rdacca.hp)

#sample data in ade4 package
require(ade4)
```

```r
data(doubs)
#fish species as response variables
spe <- doubs$fish
#environmental factors as explanatory variables
env <- doubs$env
#remove empty site 8 without species
spe <- spe[-8,]
env <- env[-8,]
#'dfs' is a variable containing locations rather than environmental
#factor. We also removed it from the 'env' data frame
env <- env[, -1]
#the usual Hellinger-transformation of the species dataset
spe.hel <- decostand(spe, "hellinger")
#selecting variables via ordistep() in vegan
ordistep(rda(spe.hel~.,env))
```

```r
#three selected variables: alt,oxy and bdo
rdacca.hp(spe.hel,env[,c("alt","oxy","bdo")], method="RDA", type = "adjR2",trace = TRUE, plot.perc = FALSE)
```

```
## $Method_Type
## [1] "RDA"   "adjR2"
##
## $R.squared
## [1] 0.5402
##
## $Var.part
##                              Fractions   % Total
## Unique to alt                   0.1942     35.96
## Unique to oxy                   0.1367     25.31
## Unique to bdo                   0.0871     16.13
## Common to alt, and oxy           0.0467      8.64
## Common to alt, and bdo          -0.0179     -3.31
## Common to oxy, and bdo           0.0131      2.43
## Common to alt, oxy, and bdo     0.0801     14.84
## Total                            0.5402    100.00
##
## $Hier.part
##     Independent I.perc(%)
## alt      0.2354     43.58
## oxy      0.1933     35.78
## bdo      0.1115     20.64
```

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

$Method_Type: This bit shows the type of canonical analysis and whether the original or adjusted R^2 were used in the analysis. In the example, it is `RDA` and `adjR2`.

$R.squared: `type="adjR2"`. In this example, the amount of variation of the fish data matrix explained by the three predictors (**alt**, **oxy** and **bdo**) is 54.02%.

$Var.part: It contains output that lists with R^2adj values for all fractions based on variation partitioning (commonality analysis) if setting `trace=TRUE`. In this example, with three predictors, commonality analysis decomposes R^2adj into all seven fractions (i.e., 2^3-1) of unique and common effects. Fractions can be either positive or negative (Nimon & Reio 2011, Ray-Mukherjee *et al* 2014). In this case, the common effect between *alt* and *bod* is negative (-0.0179). Negative common (shared) variation is possible when predictors act as suppressors of other predictors (Pedhazur 1997; Nimon & Reio 2011; Ray-Mukherjee *et a*l 2014). All fractions (except the residual) sum to the total R^2adj (i.e., the value of $R.squared).

`$Hier.part:` This bit is a matrix containing the independent contribution of each predictor in the "Independent" column, and their percentage in the "I.perc" column. In this case, the independent contributions of *alt*, *oxy* and *bdo* are 0.2354, 0.1933 and 0.1115, respectively; note again that they sum to the overall R^2adj (0.5402). One easily know how to calculate the independent contribution based the result of variation partitioning. For instance, the value (0.1933) for *oxy* is the sum of its unique effect and three average common effects (i.e., 0.1367+ 0.0467/2 +0.0131/2 + 0.0801/3). `rdacca.hp` also produces a bar graph based on the ggplot2 package (Wickham 2016) (Fig. 2). Users are encouraged to use other graph functions to generate other types of plots based on the numerical output of `rdacca.hp`.

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

```
rdacca.hp(spe.hel,list(envtopo=env[,1:3],envchem=env[,4:10]), method
="RDA", type = "adjR2",trace = TRUE, plot.perc = TRUE)

## $Method_Type
## [1] "RDA"    "adjR2"
##
## $R.squared
## [1] 0.5558
##
## $Var.part
##                               Fractions   % Total
## Unique to envtopo                 0.0814     14.64
## Unique to envchem                 0.2290     41.21
## Common to envtopo, and envchem    0.2454     44.15
## Total                             0.5558    100.00
##
## $Hier.part
##         Independent I.perc(%)
## envtopo      0.2041     36.72
## envchem      0.3517     63.28
```

Discussion

Multivariate regressions and canonical analysis are quintessential quantitative frameworks to tackle many ecological and evolutionary problems. Variation partitioning and hierarchical partitioning (HP) are frameworks that allow going beyond the usual and standard interpretation of partial slopes (see Ray-Mukherjee *et al.* 2014 for a review). That said, our presentation of HP allows realizing that both methods are interrelated, and that variation partitioning can serve as a precursor of HP. While variation partitioning estimates unique and common variation in a single full model containing all variables, HP is based on principles of all subset regression and model averaging which are known to improve model inference and interpretability over traditional model selection procedures (Burnham & Anderson 2002). The two main motivations for using variation partitioning and HP over traditional model selection procedures are that: a) candidate predictors are conditional on the variance explained by the models already retained by the selection procedure (Thompson 1995; Nathans, Oswald & Nimon 2012). As such, a predictor that can be quite relevant in multiple sub-models (indicating overall importance) may not be retained by model selection procedures; b) model selection can be heavily influenced by sampling variation (Thompson 1995; Nathans, Oswald & Nimon 2012); if another sample were to be used, the selected predictors could vary. By using HP, one can analyze variable importance over all possible models.

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

Currently, our `rdacca.hp` package has been used in peer-reviewed papers (e.g. Li *et al.* 2020; Song *et al.* 2020; Sun *et al.* 2020; Wang *et al.* 2020; Xiong *et al.* 2020; Zhou *et al.* 2020). Researchers using the `rdacca.hp` package in their studies, should cite this article and the `rdacca.hp` package as well. Citation information can be obtained by typing: `citation("rdacca.hp")`.

Acknowledgements

The research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19050404) and the National Science and Technology Basic Resources Survey Program of China (2019FY100204). PP-N was supported by the Canada Research Chair (CRC) program. Otherwise there is no conflict of interest among the authors.

Authors' contributions

JSL conceived the idea. JSL and PP-N wrote the package and conducted the analysis. All authors participated in writing multiples drafts.

References

Bi, J. (2012) A review of statistical methods for determination of relative importance of correlated predictors and Identification of drivers of consumer liking. *Journal of Sensory Studies,* **27,** 87-101.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology,* **73,** 1045-1055.

Budescu, D.V. (1993) Dominance analysis - a new approach to the problem of relative importance of predictors in multiple-regression. *Psychological Bulletin,* **114,** 542-551.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference : a practical information-theoretic approach,* second edn. Springer-Verlag, New York.

Chevan, A. & Sutherland, M. (1991) Hierarchical partitioning. *American Statistician,* **45,** 90-96.

Cox, L.A. (1985) A new measure of attributable risk for public health applications. *Management Science,* **31,** 800-813.

Ezekiel, M. (1930) *Methods of Correlational Analysis*. Wiley, New York

Grömping, U. (2006) Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software,* **17,** 1-27..

Grömping, U. (2007) Relative importance in linear regression based on variance decomposition. *American Statistician,* **61,** 139-147.

Grömping, U. (2009) Variable importance assessment in regression: Linear regression versus random forest. *American Statistician,* **63,** 308-319.

Grömping, U. (2015) Variable importance in regression models. *WIREs Computational Statistics,* **7,** 137-152.

Kruskal, W. (1987) Relative importance by averaging over orderings. *American Statistician,* **41,** 6-10.

Legendre, P. & Anderson, M.J. (1999) Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs,* **69,** 1-24.

Legendre, P. & Legendre, L. (2012) *Numercial Ecology,* Third edn. Elsevier.

Li, J.Y., Wang, N.L., Dodson, J., Yan, H., Zhang, X.J., Jia, P.W. & Seppa, H. (2020) Holocene negative coupling of summer temperature and moisture availability over southeastern arid Central Asia. *Climate Dynamics,* **55,** 1187-1208.

Lindeman, R.H., Merenda, P.F. & Gold, R.Z. (1980) *Introduction to Bivariate and Multivariate*

Analysis. Scott Foresman, Glenview, IL.

Mac Nally, R. & Walsh, C.J. (2004) Hierarchical partitioning public-domain software. *Biodiversity and Conservation,* **13,** 659-660.

Nathans, L.L., Oswald, F.L. & Nimon, K. (2012) Multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research & Evaluation,* **17,** 1-19.

Navarrete, C.B. & Soares, F.C. (2020) dominanceanalysis: Dominance Analysis. R package version 2.0.0.

Newton, R.G. & Spurell, D.J. (1967) Examples of the use of elements for classifying regression analysis. *Applied Statistics,* **16,** 165-172.

Nimon, K. & Reio, T. (2011) Regression commonality analysis: a technique for quantitative theory building. *Human Resource Development*, **10**,329– 340.

Nimon, K., Oswald, F.L. & Roberts, J.K. (2013) Yhat: Interpreting regression effects. R package version 2.0.0.

Økland, R.H. (2003) Partitioning the variation in a plot-by-species data matrix that is related to n sets of explanatory variables. *Journal of Vegetation Science,* **14,** 693-700.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2019) vegan: Community Ecology Package. R package version 2.5-4.

Oksanen, J., Kindt, R., Legendre, P. & O'Hara, R.B. (2006) vegan: Community Ecology Package. R package version 1.8-1.

Pedhazur, E.J. (1997) Multiple Regression in Behavioral Research: Explanation and Prediction, 3rd edn. Harcourt Brace, Orlando, FL.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology,* **87,** 2614-2625.

R Development Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhyā A,* **26,** 329-358.

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D.W., Slotow, R. & Hamer, M. (2014) Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution,* **5,** 320-328.

Song, S.S., Zhang, C., Gao, Y., Zhu, X.Y., Wang, R.H., Wang, M.D., Zheng, Y.L., Hou, L.J., Liu, M. & Wu, D.M. (2020) Responses of wetland soil bacterial community and edaphic factors to two-year experimental warming and Spartina alterniflora invasion in Chongming Island. *Journal of Cleaner Production,* **250**, doi.org/10.1016/j.jclepro.2019.119502.

Soofi, E.S. (1992) A generalizable formulation of conditional logit with diagnostics. *Journal of the American Statistical Association,* **87,** 812-816.

Sun, X.J., Diekman, M., Yan, X., Zou, Y., Sang, W.G. & Axmacher, J.C. (2020) Diversity and seasonal changes in carabid assemblages of a mature, secondary and plantation forest mosaic in the Zhangguangcai Mountains in northeastern China. *Insect Conservation and Diversity,* **13,** 340-350.

ter Braak, C.J.F. (1986) Canonical correspondence-analysis - a new eigenvector technique for multivariate direct gradient analysis. *Ecology,* **67,** 1167-1179.

Thioulouse, J., Dray, S., Dufour, A.B., Siberchicot, A., Jombart, T. & Pavoine, S. (2018) *Multivariate*

Analysis of Ecological Data with ade4. Springer.

Thompson, B. (1995) Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement,* **55,** 525-534.

Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse d'état, Besançon.

Verneaux, J., Schmitt, A., Verneaux, V. & Prouteau, C. (2003) Benthic insects and fish of the Doubs River system: typological traits and the development of a species continuum in a theoretically extrapolated watercourse. *Hydrobiologia,* **490,** 63-74.

Walsh, C.J. & Mac Nally, R. (2013) hier.part: Hierarchical Partitioning. R package version 1.0-4.

Walsh, C.J., Papas, P.J., Crowther, D. & Yoo, J. (2004) Stormwater drainage pipes as a threat to a stream-dwelling amphipod of conservation significance, Austrogammarus australis, in southeastern Australia. *Biodiversity and Conservation,* **13,** 781-793.

Wang, Y., Yang, X.D., Ali, A., Lv, G.H., Long, Y.X., Wang, Y.Y., Ma, Y.G. & Xu, C.C. (2020) Flowering phenology shifts in response to functional traits, growth form, and phylogeny of woody species in a desert area. *Frontiers in Plant Science,* **11**, doi: 10.3389/fpls.2020.00536

Wickham, H. (2016) *Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Xiong, Q.L., Luo, X.J., Liang, P.H., Xiao, Y., Xiao, Q., Sun, H., Pan, K.W., Wang, L.X., Li, L.J. & Pang, X.Y. (2020) Fire from policy, human interventions, or biophysical factors? Temporal-spatial patterns of forest fire in southwestern China. *Forest Ecology and Management,* **474,** doi.org/10.1016/j.foreco.2020.118381

Zhou, S.L., Sun, Y., Li, Z.X. & Huang, T.L. (2020) Characteristics and driving factors of the aerobic denitrifying microbial community in Baiyangdian Lake, Xiong'an New Area. *Microorganisms,*

8, doi:10.3390/microorganisms8050714

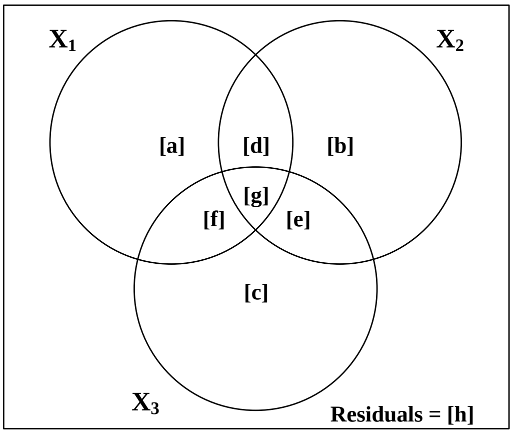

Figure 1. Venn diagram representing the variation partitioning of a response matrix **Y** regressed against three correlated predictors (or groups of predictors as in variation partitioning). All fractions add to 100% and the bounding rectangle represents the total variation in the response data (i.e.,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

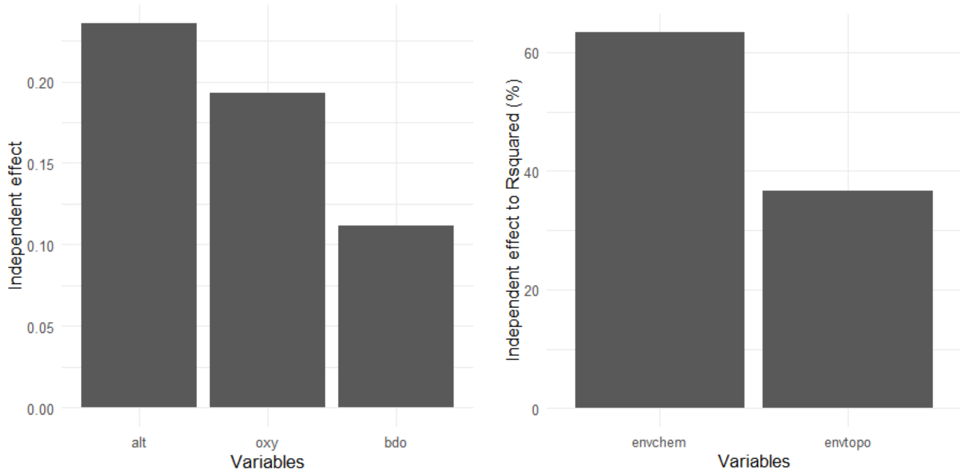

Figure 2. The graphic output of `rdacca.hp` function for the Doubs River fish data. left panel: original independent contribution for three predictors (`plot.perc=FLASE`) and right panel: their relative contribution for two groups to the overall adjusted R^2 (`plot.perc=TRUE`) .