

R程序包“rdacca.hp”在生态学数据分析中的应用：案例与进展

刘尧^{1,2} 于馨^{3,4} 于洋⁵ 胡文浩⁶ 赖江山^{1,3,7*}

¹南京林业大学生物与环境学院, 南京 210037; ²厦门大学环境与生态学院, 福建厦门 361102; ³中国科学院植物研究所植被与环境变化国家重点实验室, 北京 100093; ⁴中国科学院大学, 北京 100049; ⁵北京林业大学水土保持学院, 北京 100083; ⁶浙江农林大学风景园林与建筑学院, 杭州 311300; ⁷南京林业大学数量生态学研究中心, 南京 210037

摘要 定量评估不同变量对群落组成的贡献是群落生态学分析的热点问题。但在具体的分析情景中, 因子间的共线性与解释率的重叠对评估不同因子重要性造成了较大困难。基于这一问题, R程序包“rdacca.hp”通过引入层次分割法(HP)的理念, 在所有可能的模型子集下为各解释变量(或解释变量组)分配单独效应, 为典范分析中共线性解释变量的相对重要性评估提供了新的定量指标。目前, “rdacca.hp”包已经成为群落生态学分析的重要工具。为进一步促进用户对“rdacca.hp”包的理解与运用, 该文通过引入一个分析塑造甲螨(Oribatida)群落的重要环境和空间驱动因素的实例, 重点展示了使用该程序包进行典范分析的一般步骤。随后对近期应用“rdacca.hp”包开展分析的相关研究进行文献计量学分析, 结果表明该程序包自上线以来已被广泛用作解决生态学、环境科学及相关学科问题的基本定量框架。最后, 该文对“rdacca.hp”包的未来应用和升级进行了展望。总之, 该文旨在使国内学者们进一步加深对“rdacca.hp”包的认识和应用。

关键词 “rdacca.hp”包; 典范分析; 相对重要性; 层次分割; 群落组成

刘尧, 于馨, 于洋, 胡文浩, 赖江山 (2023). R程序包“rdacca.hp”在生态学数据分析中的应用: 案例与进展. 植物生态学报, 47, 134-144. DOI: 10.17521/cjpe.2022.0314

Application of “rdacca.hp” R package in ecological data analysis: case and progress

LIU Yao^{1,2}, YU Xin^{3,4}, YU Yang⁵, HU Wen-Hao⁶, and LAI Jiang-Shan^{1,3,7*}

¹College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China; ²School of Environment and Ecology, Xiamen University, Xiamen, Fujian 361102, China; ³State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; ⁴University of Chinese Academy of Sciences, Beijing 100049, China; ⁵School of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China; ⁶College of Landscape Architecture, Zhejiang A&F University, Hangzhou 311300, China; and ⁷Research Center of Quantitative Ecology, Nanjing Forestry University, Nanjing 210037, China

Abstract

Quantitative estimation of the contribution of predictory variables to community composition is a hotspot in community ecology. However, multicollinearity and joint contributions among predictors make it difficult to estimate the importance of predictor in specific analysis scenarios. To address this issue, the “rdacca.hp” package provides a new quantitative indicator by introducing the concept of hierarchical partitioning (HP) to assign individual effects for individual predictors (or groups of predictors) across all possible model subsets. The package solves the problem of estimating the relative importance of predictors with multicollinearity in canonical analysis. The “rdacca.hp” package has become an important tool for community ecological analysis. To further promote users’ understanding and use of the “rdacca.hp” package, we demonstrate the general steps for using this package in canonical analysis with an example analyzing the important environmental and spatial drivers that shape the oribatid mites (Oribatida) community. Subsequently, we conduct a bibliometric analysis of recent studies using “rdacca.hp” package. The results show that, since its launch, the package has been widely used as a fundamental quantitative framework in ecology, environmental science and related disciplines. Finally, we discuss the further application and extension of the “rdacca.hp” package. In conclusion, this paper aims to advocate the understanding and application of the “rdacca.hp” package for domestic researchers.

Key words “rdacca.hp” package; canonical analysis; relative importance; hierarchical partitioning; community composition

收稿日期Received: 2022-07-26 接受日期Accepted: 2022-09-28

基金项目: 国家自然科学基金(32271551). Supported by the National Natural Science Foundation of China (32271551).

* 通信作者Corresponding author (lai@njfu.edu.cn)

Liu Y, Yu X, Yu Y, Hu WH, Lai JS (2023). Application of “rdacca.hp” R package in ecological data analysis: case and progress. *Chinese Journal of Plant Ecology*, 47, 134-144. DOI: 10.17521/cjpe.2022.0314

探索多响应变量与多解释变量之间关系的约束性排序类方法(或称典范分析, canonical analysis)已被广泛用作生态学数据分析的基本框架, 具体包括冗余分析(RDA)、典范对应分析(CCA)和基于距离的冗余分析(db-RDA)等(Legendre & Legendre, 2012)。在群落生态学研究, 典范分析的主要目标之一是定量给定环境因子(解释变量)对目标生物群落组成(响应变量)的解释程度(即 R^2), 以揭示物种组成(多样性)的环境响应机制。但除了模型的总体解释率之外, 解释变量的相对重要性也是研究人员普遍感兴趣的问题(Peres-Neto *et al.*, 2006)。解释变量的相对重要性的一种常用判断依据是每个变量对 R^2 的贡献比例(Johnson & LeBreton, 2004)。当所有解释变量彼此完全相互独立(即相关系数为0), 则不同解释因子的单独 R^2 之和就等于总 R^2 , 这种情况被称为 R^2 可加性(Genizi, 1993)。但这种情况在真实场景下很难出现, 由于生态学数据多来自野外观测, 不同指标间的相互关系错综复杂, 存在共线性是普遍现象(Graham, 2003)。以土壤理化数据为例, 土壤有机质含量、总氮含量、pH等指标通常具有相关性, 如将这些变量全部纳入模型中, 就会出现解释变量的共线性。共线性对解释率的影响主要体现在不同因子的解释变差会存在共同解释的部分, 统计上被称为共享 R^2 (Legendre & Legendre, 2012)。共享 R^2 的比例越大, 说明共线性越强。由于共享 R^2 的存在, 导致不

同环境因子的简单 R^2 (即单个解释变量作为唯一解释变量时模型的 R^2)之和通常会大于模型的总体 R^2 , 因此, 有必要对共享 R^2 进行分割, 以明确不同解释变量对响应变量的独立贡献。

分割共享 R^2 的难点在于很难从数学上确定共享 R^2 由相关变量中的哪个引起, 因此就很难将共享的 R^2 准确分配给不同的变量。面对这一问题, 生态学家通常采用模型筛选、变量顺序 R^2 与变差分解3种方法进行处理。其中, 模型筛选主要通过减少模型中的共线性变量来降低模型的共线性(Burnhan & Anderson, 2004; Blanchet *et al.*, 2008)。这一方法大多基于一定的准则(如 p 值、赤池信息准则(AIC)值)得到一个所谓的最优模型, 但这种仅通过统计学手段得到的最优变量集合可能不具生态学意义, 且仍可能存在共线性的问题。另一种替代方案是人为筛选, 依据主成分分析(PCA)、方差膨胀因子(VIF)等方法评估解释变量间的相关关系或共线性强弱, 并综合考虑变量的身份以及其生物学意义, 最终人为做出决策。但这种筛选变量的过程无疑带有主观判断, 很难让人信服。变量顺序 R^2 的方法是将解释变量按一定的顺序加入模型中, 并观察 R^2 的增加情况来判断不同变量的贡献(图1A)。通常情况下, 按照 p 值由小到大或增加的 R^2 由大到小的顺序进行模型添加, 加入变量前后 R^2 的差值就被称为随机 R^2 , 或条件效应(conditional effect)(Šmilauer & Lepš, 2014)。

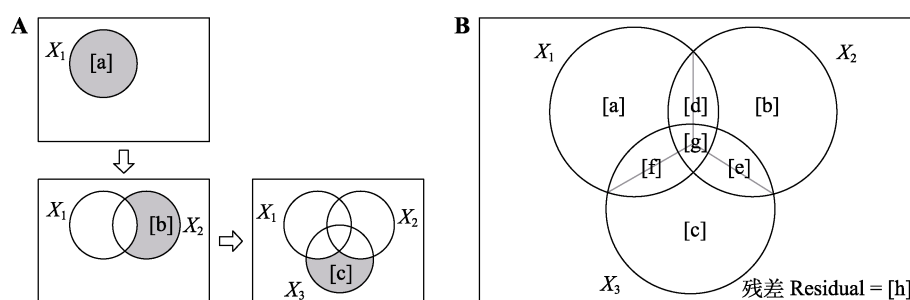


图1 以3组解释变量 X_1 、 X_2 和 X_3 为例, 使用韦恩图概括了响应变量 Y 的总变差组成。**A**, 基于变量的顺序 R^2 的方法, 根据分配的顺序, [a]、[b]和[c]分别为 X_1 、 X_2 和 X_3 的条件效应。**B**, 基于变差分解和层次分割法, [a]、[b]和[c]分别为 X_1 、 X_2 和 X_3 的边际效应, [d]、[e]和[f]分别为 X_1 、 X_2 和 X_3 两两之间的共同效应, [g]为三者之间的共同效应, X_1 、 X_2 和 X_3 的单独效应分别为[a] + [d]/2 + [f]/2 + [g]/3、[b] + [d]/2 + [e]/2 + [g]/3以及[c] + [e]/2 + [f]/2 + [g]/3。残差[h]代表了未被 X_1 、 X_2 和 X_3 解释的 Y 的变差部分。

Fig. 1 Venn diagram representing the variation composition of a response matrix Y regressed against three correlated predictors of X_1 , X_2 , and X_3 . **A**, According to the order of assignment, [a], [b] and [c] are the conditional effect of X_1 , X_2 and X_3 , respectively. **B**, Based on variation partitioning and hierarchical partitioning, [a], [b] and [c] are the marginal effect of X_1 , X_2 and X_3 , respectively; [d], [e] and [f] are the common effect among X_1 , X_2 and X_3 , respectively; [g] is the common effect among X_1 , X_2 and X_3 ; the individual effect of X_1 , X_2 and X_3 can be expressed as [a] + [d]/2 + [f]/2 + [g]/3, [b] + [d]/2 + [e]/2 + [g]/3 and [c] + [e]/2 + [f]/2 + [g]/3. Residual [h] represents the fraction of Y that is not explained by X_1 , X_2 or X_3 .

CANOCO软件与R程序包“adespatial”的forward.sel函数均采用了这种方法,在过去的研究中被广泛使用。这种方法的问题在于解释变量所分配的顺序 R^2 显然取决于变量进入模型的顺序。如果解释变量存在共线性,共享 R^2 更容易分配给先加入模型的解释变量,而后加入的解释变量所分配的共享 R^2 变少,导致先加入模型的解释变量占更大的优势。除非分析者具有极强的先验知识,这种方法实际上存在一定问题。第三种方法是变差分解 (variation partitioning, VP),也常称作方差分解,最早出现于20世纪60年代用于多元回归的共性分析 (commonality analysis)(Newton & Spurrell, 1967; Mood, 1969),20世纪90年代被引入至典范分析中 (Borcard *et al.*, 1992; Peres-Neto *et al.*, 2006)。VP可以将模型的 R^2 分解为由每个变量(或变量组)所能独自解释的组分(称为独自效应(unique effect),或边际效应(marginal effect)以及共同解释的组分(称为共同效应(common effect))来理解解释变量之间复杂的关系(图1B)。然而随着纳入的变量数目的增加将导致计算量呈指数型增长,因此VP分组数量经常受到一定限制,例如“vegan”包的varpart函数最多只能纳入4组变量。此外,即使不对分组进行限制,也依然无法解决共同效应的分割问题。

事实上,“共线性解释变量相对重要性的比较”在多元回归领域一直是活跃的主题。20世纪80年代,由Lindeman等(1980)、Cox Jr (1985)和Kruskal (1987)分别独立地提出了一种“平均排序(averaging over orderings)” (也称LMG指数)的方法,基于所有模型子集获取解释变量的相对重要性。与Chevan和Sutherland (1991)提出的层次分割 (hierarchical partitioning, HP)结果一样,并推广至广义线性模型 (GLM)。R程序包“relaimpo”可用于计算多元回归中解释变量的LMG值,并提供自举法检验LMG值的

显著性(Grömping, 2007)。“hier.part”包基于HP来估计GLM中解释变量的相对重要性(Mac Nally & Walsh, 2004)。最近, Lai等(2022a)将VP和HP建立数学联系,首次将HP算法引入典范分析并开发了“rdacca.hp”包来获取单个典范分析解释变量的解释率,同时实现了不限解释变量数量限制的VP。HP强调基于所有可能的模型子集下各解释变量的单独效应(individual effect),等同于解释变量的边际效应加上与其他解释变量的共同效应的平均分配值(图1B),为典范分析中评估共线性的解释变量的相对重要性提供了新型定量指标。

为进一步推广该方法在生态学领域的应用,帮助科学工作者在复杂环境背景下寻找关键因子,本文旨在: (1)介绍“rdacca.hp”包的主要功能; (2)通过案例分析详解“rdacca.hp”包的使用; (3)结合“rdacca.hp”包自发布以来的相关进展对该程序包在生态学及更多学科领域中的应用和升级进行展望。

1 “rdacca.hp”包的结构

截至目前,“rdacca.hp”包的最新版本为1.0-8 (发布时间是2022年4月5日)。用户可以在R语言中输入“install.packages(“rdacca.hp”)”直接从CRAN安装“rdacca.hp”包,或者使用“devtools::install_github(“laijiangshan/rdacca.hp”)”在GitHub获取。该程序包封装的函数及核心功能见表1,各函数具体参数和输出的解释可进一步参阅R文档。

2 “rdacca.hp”包的操作案例——塑造甲螨 (Oribatida)群落的重要环境和空间驱动因素分析

下面使用“vegan”包附带的甲螨数据集演示“rdacca.hp”包的使用。该案例数据由蒙特利尔大学劳伦蒂斯生态研究站于1989年在Lac Geai湖边的泥

表1 “rdacca.hp”包的函数介绍
Table 1 Function of “rdacca.hp” package

函数 Function	介绍 Description
rdacca.hp	运行不限解释变量数量的变差分解和层次分割,计算每组解释变量对典范分析中解释变异(R^2 或校正 R^2)的边际、共同和单独效应 of predictors. Output include the marginal, common, as well as individual effect of each (group) variables to total R^2 or adjusted R^2
permu.hp	对解释变量的单独效应进行置换检验以评估其显著性 Implements the significance testing for individual contribution from hierarchical partitioning by permutation routine
plot.rdaccahp	基于“rdacca.hp”函数的输出,绘制柱形图展示解释变量单独效应 Plot a bar plot of the individual contribution of predictors based on the output of rdacca.hp()

表2 R语言“vegan”包的甲螨多度及环境和空间因子数据集

Table 2 Oribatid mites abundance dataset with environmental and spatial variables in R package “vegan”

数据框 Data frame	介绍 Description
mite	物种多度矩阵, 包含70个采样点的共计35种甲螨形态种的个体计数 Species abundance matrix that contains the abundance of 35 oribatid mites morphospecies at 70 sampling sites
mite.env	环境因子矩阵, 包含70个采样点的共计5种环境变量: 基质密度、基质含水量、基质类型、灌丛密度和微地形特征 Environmental matrix that contains 5 environmental variables at 70 sampling sites, including substrate density, water content of the substrate, substrate type, shrub density, and microtopography
mite.pcnm	空间因子矩阵, 基于70个采样点的地理坐标, 通过邻体矩阵主坐标分析(PCNM)构建的特征向量, 反映了70个采样点的潜在空间结构 Spatial matrix, which is based on the geographic coordinates of 70 sampling sites and constructed by principal coordinates of neighbour matrices (PCNM) (Borcard & Legendre, 2002), represents the potential spatial structure of 70 sampling points

炭藓地采集的70个土壤样本构成, 调查甲螨多样性并记录了采样点的环境特征和空间位置信息(表2)(Borcard *et al.*, 1992)。下文分析的一个主要目的是揭示该区域生物群落构建的关键环境或空间驱动因素。

```
> data(mite, package="vegan")
> data(mite.env, package="vegan")
> data(mite.pcnm, package="vegan")
```

2.1 单个解释变量的相对重要性分析

我们首先通过典范分析确定环境异质性和甲螨群落组成的影响。考虑到物种多度矩阵中存在较多0值, 预先进行Hellinger转化以消除极端值的影响并减少高丰度物种的权重(Legendre & Gallagher, 2001)。随后, 基于RDA的典范分析用于拟合物种多度(作为响应变量)与所有5个环境因子(作为解释变量)的多变量回归, 并通过VP和HP为各环境因子分配相应的重要性指标。未对环境因子进行共线性评估或者其他任何方式的筛选, 目的是确保获取所有可能的模型子集, 同时比较了VP和HP方法的差异。

```
> library(rdacca.hp)
> mite.hel <- decostand(mite, method="hellinger")
> mod <- rdacca.hp(mite.hel, mite.env, method="RDA", var.part=TRUE, type="adjR2")
> UpSetVP::upset_vp(mod, plot.hp=TRUE, order.part="degree")
> permu.hp(mite.hel, mite.env, method="RDA", type="adjR2", permutations=999)
```

RDA显示5个环境因子共解释了物种多度总变差的43.7% (校正 $R^2 = 0.437$) (图2)。“rdacca.hp”包成功实现了大于4组解释变量的VP, 然而随着解释变量数量的增加, 从中解读所有边际效应和共同效应来理解变量间的复杂关系非常具有挑战性。相比之下, HP通过继续分解并组合这些组分来分配单独效

应, 提供了更便捷的模型解读框架: 5个环境因子的相对重要性排序依次为基质含水量(16.42%)、灌丛密度(10.18%)、微地形特征(8.43%)、基质类型(5.43%)和基质密度(3.2%)。所有环境因子的单独效应均显著, 但必须注意的是, 由于解释变量的相对重要性通常用作探索性框架而不是推理工具, 置换检验的结果仅用作参考(Lai *et al.*, 2022a)。尽管过去流行使用VP的边际效应来估计解释变量的相对重要性, 但片面的信息容易产生误报。例如, 仅比较边际效应时, 灌丛密度(2.88%)则是对甲螨群落组成贡献最小的环境因子。两种方法产生分歧的一个主要原因在于灌丛密度和基质含水量以及微地形特征等其他环境因子间存在较大的共同效应(归因于共线性), 后者忽略这些共同解释的部分导致灌丛密度的贡献被严重低估。此外, 出于解释变量之间的抑制效应或校正 R^2 等原因, 部分共同效应出现负值。观察到基质密度的边际效应(3.63%) > 单独效应(3.2%), 意味着仅关注边际效应也会使个别解释变量的重要性被高估。综上, HP在揭示变量的混杂关系中具有优势。

2.2 解释变量组的相对重要性分析

“rdacca.hp”包同样提供了用于解释变量组的VP和HP。我们继续将空间因子纳入典范分析, 使用RDA的VP和HP量化空间结构和环境异质性对甲螨群落组成的贡献。

```
> iv <- list(PCNM=mite.pcnm, ENV=mite.env)
> mod <- rdacca.hp(mite.hel, iv, method="RDA", var.part=TRUE, type="adjR2")
> UpSetVP::upset_vp(mod, plot.hp=TRUE, order.part="degree")
> permu.hp(mite.hel, iv, method="RDA", type="adjR2", permutations=999)
```

结果显示, 空间因子(单独效应27.39%, $p <$

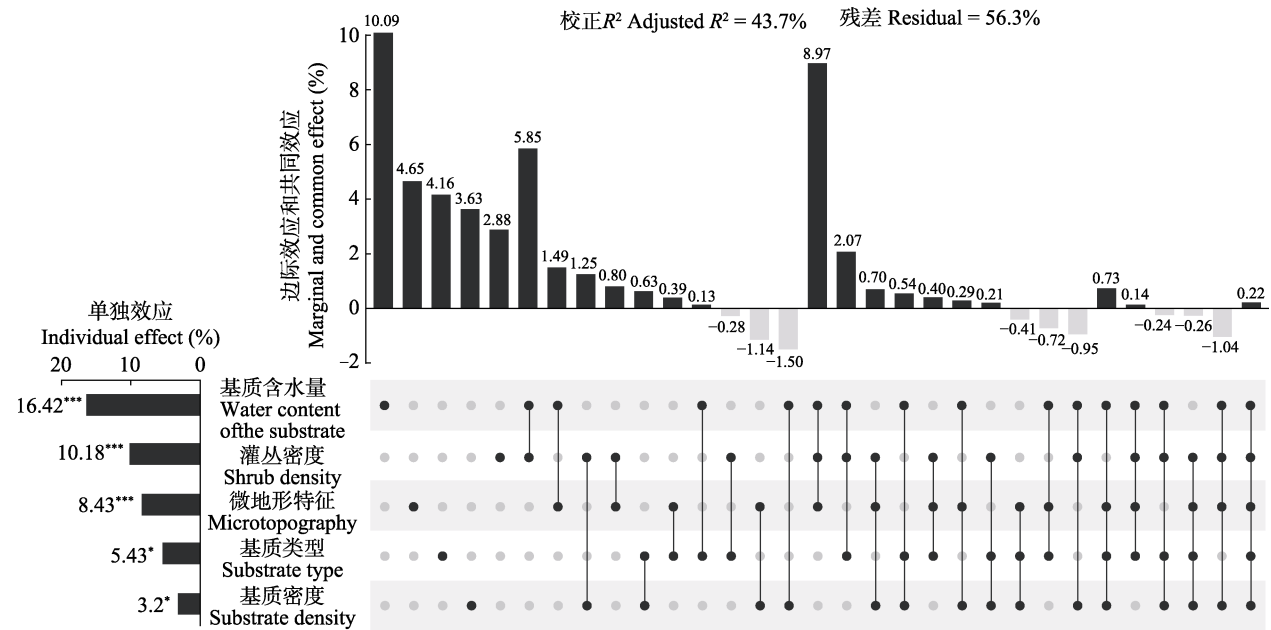


图2 5种环境因子对甲螨物种多度影响的相对重要性。变差分解和层次分割分析结果使用UpSet图来呈现。右侧点阵图中, 每行对应一个环境因子。对于每一列, 孤立黑点表示各环境因子的边际效应, 多点间连线表示这些环境因子间的共同效应, 各组分解释的变差百分比(来自变差分解)展示在上方柱形图中。左侧柱形图为各环境因子的单独效应(来自层次分割), 其值等同于该环境因子的边际效应加上与其他环境因子的共同效应的平均分配值。*, $p < 0.05$; ***, $p < 0.001$ 。

Fig. 2 UpSet matrix layout of variation partitioning and hierarchical partitioning results to show the relative importance of five environmental factors on oribatid mites abundance. In the dot-matrix plot on the right, each row corresponds to an environmental factor. For each column, the isolated black dot represents the marginal effect of each environmental factor, lines connecting multiple dots represent the common effect among these corresponding environmental factors, and the percentage of variation explained by each component (from variation partitioning) are shown in the top column diagram. Column diagram on the left shows individual effect of each environmental factor (from hierarchical partitioning), its value is equal to its marginal effect plus its average shared common effect with other environmental factors. *, $p < 0.05$; ***, $p < 0.001$.

0.001)和环境因子(单独效应26.41%, $p < 0.001$)均显著影响甲螨群落组成, 二者共同解释了物种多度总变差的53.8% (校正 $R^2 = 0.538$) (图3)。由于相差不足1%, 暗示扩散限制和环境选择压力在塑造当地甲螨群落中的重要性近似相等。值得注意的是两组解释变量存在较高比例的共同效应(34.53%), 可能意味着这些环境因子具有空间结构性, 促使群落形成类似的空间格局, 或者其他空间结构过程(如历史事件或生物相互作用)引发的环境和群落的一致响应。

3 “rdacca.hp”包的相关应用实例

最后, 我们汇总了使用“rdacca.hp”包开展分析的相关研究, 以进一步加深用户对该程序包应用场景的认识。以Google学术为数据源检索关键词“rdacca.hp”和“rdaenvpart” (“rdacca.hp”包的前身, 下文统称为“rdacca.hp”)并根据内容进行甄别, 截至2022年8月31日共计93篇正式发表的论文。“rdacca.hp”包自发布于CRAN以来便得到学术界的

高度关注, 被引次数呈现逐年递增趋势, 特别是2022年1月该方法学论文(Lai *et al.*, 2022a)在线发表后被更多国内外学者所熟知(图4)。继续梳理相关研究涉及的学科领域、研究区域或环境类型、典范分析的响应变量和解释变量类别的关键词并进行共词分析, 结果同样显示“rdacca.hp”包正在多学科交叉的数据分析中流行(图5)。其中, 环境科学与生态学、生物学和农林科学是“rdacca.hp”包集中出现的领域, 并涵盖陆地、水域的多种自然系统以及耕地、城市等人工系统, 部分面向微观系统或上升到景观尺度。考虑到“rdacca.hp”包的主要目标是提供定量指标量化响应变量和解释变量的关系, 下文依据这些研究中典范分析的对象和目的, 举例讨论该程序包的当前应用。

3.1 生物多样性

在所有典范分析的对象中, 生物多样性、群落结构及其与环境、气候、地形或其他生物因子的关系受到最多关注, 陆地生物群落的研究比水生生物群落更广泛(图5, 图6)。揭示生物多样性和群落构建

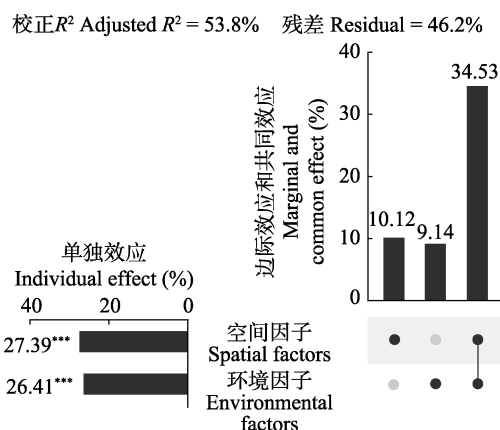


图3 空间和环境因子对甲螨物种多度影响的相对重要性。变差分解和层次分割分析结果使用UpSet图来呈现。右侧点阵图中, 3列分别表示空间因子和环境因子的边际效应以及二者的共同效应(来自变差分解)。左侧柱形图表示二者的单独效应(来自层次分割)。***, $p < 0.001$ 。

Fig. 3 UpSet matrix layout of variation partitioning and hierarchical partitioning results to show the relative importance of spatial and environmental factors on oribatid mites abundance. In the dot-matrix plot on the right, three columns represent the marginal effect of spatial factors, the marginal effect of environmental factors, and the common effect between them, respectively (from variation partitioning). Column diagram on the left shows individual effect of spatial and environmental factors (from hierarchical partitioning). ***, $p < 0.001$.

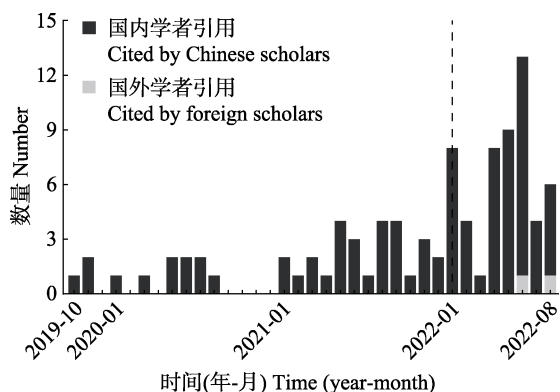


图4 引用“rdacca.hp”包的论文发表时间及数量统计。虚线表示“rdacca.hp”包相关的方法学论文的正式发表时间。

Fig. 4 Statistics of published papers citing “rdacca.hp” package (publication time and number of published paper). Dashed line indicates the publication time of methodological paper related to “rdacca.hp” package.

的驱动因素是生态学研究中的核心问题之一, 特别是在当前全球变化背景下, 对于了解生态群落和相关生态系统功能的变化至关重要(Kreft & Jetz, 2007; Hooper *et al.*, 2012; Guo *et al.*, 2018)。保护生态学更注重对相对重要性的分析, 因为明确不同生态因子的重要性对于制定保护策略具有重要价值。尽管不同生物类群对资源需求和环境适应存在异质性响应,

但多项研究的HP结果呈现出一些共性。例如, 地形或气候条件通常比土壤理化因子更主导森林和草原植被特征的变化(Liu *et al.*, 2022b; Xie *et al.*, 2022; Zhao *et al.*, 2022)。淡水湖泊浮游植物群落受养分梯度的强烈影响, 氮是关键驱动因素(Hu *et al.*, 2021; Zhang *et al.*, 2023)。这些均暗示需要采取适应措施以应对未来气候和营养负荷变化可能带来的不利影响。

虽然典范分析方法最初是为解决植物群落问题而开发, 但“rdacca.hp”包更多被应用于微生物群落分析(图6)。事实上, 很多植物生态学研究方法后来均在微生物生态学中得到验证, 这得益于近年来分子测序技术的发展推动微生物生态学研究成为热点。与天然群落相比, 与农林牧有关的土壤微生物更受青睐。在包括农田、人工林、牧场等在内的人工生态系统中, β 多样性和HP分析强调pH相较其他理化属性对土壤细菌群落的影响更直接(Mei *et al.*, 2021; Xiao *et al.*, 2021; Xu *et al.*, 2021; Li *et al.*, 2022), 而土壤真菌群落对有机碳组分的响应更敏感(Xu *et al.*, 2021; Wang *et al.*, 2022a, 2022d)。理解土地利用或管理制度如何影响土壤性质以及微生物功能, 有助于指导有益的农林牧实践并提升生态效益。同时, 植物-微生物的相互作用也被广泛关注。以根系相关微生物最具代表性, 植物种类和生理条件的重要性通常大于非生物因子(Wang *et al.*, 2022b, 2022c, 2022d)。地上植被与地下微生物群落构成维持生态功能的系统实体, 有益的生物协同作用促进养分转化并提升生产力, 为可持续生态提供了更多机会。此外, 在环境修复、发酵生产等工程领域也偶见使用“rdacca.hp”包的方法分析微生物群落构建的驱动因素, 由于相关文献数量尚少暂不讨论。

3.2 生物地球化学

作为生物与环境综合作用的结果, 生物地球化学过程塑造地质和生物历史。除了群落生态学应用, 典范分析也常用做环境中碳氮磷化学或生物产物及其驱动因素的框架(图5, 图6)。生物多样性的变化是否会导致生态系统功能和相关的生物地球化学过程的改变同样是当代生态学中的关键问题(Isbell *et al.*, 2018; van der Plas, 2019)。尽管当前案例尚少, 但仍可根据HP得知林分结构显著影响土壤微生物生物量、物种组成、酶活性以及化学计量(Xu *et al.*, 2021; Yan *et al.*, 2022a), 改善森林结构和植物多样

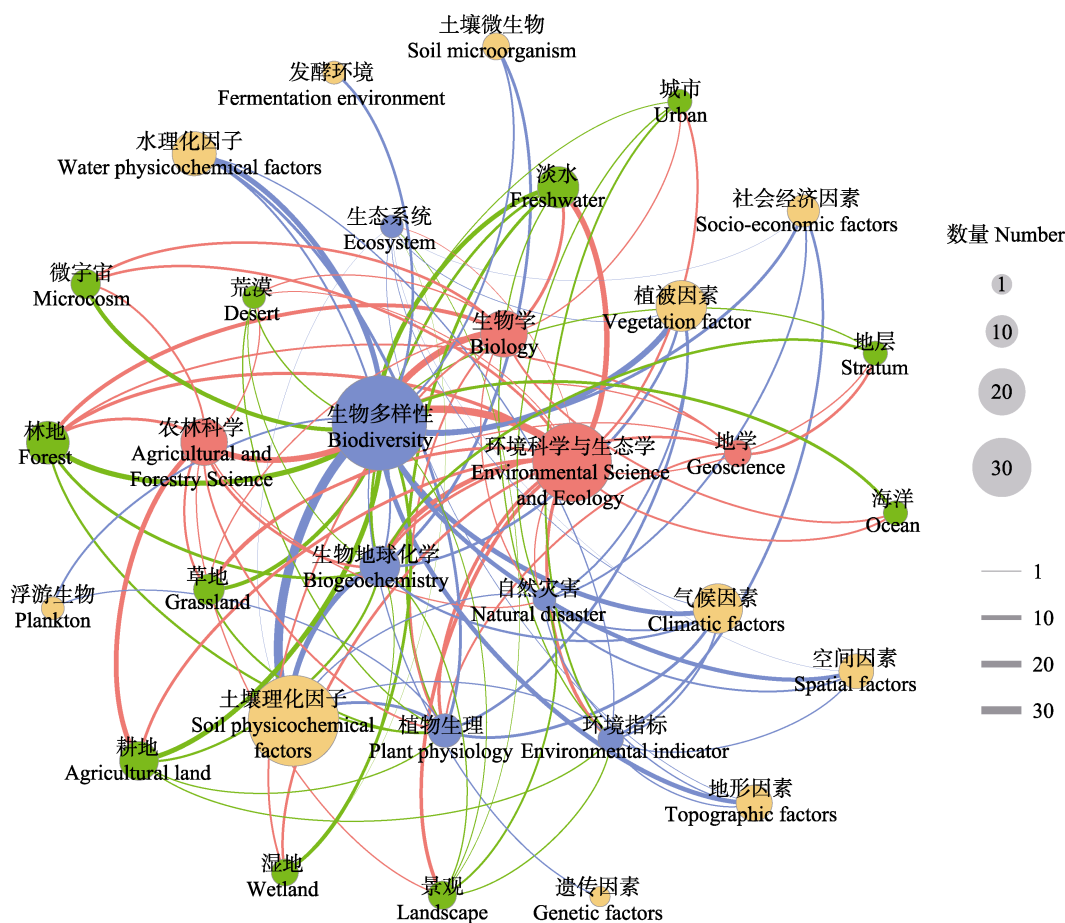


图5 引用“rdacca.hp”包的论文所属的大类学科(红色)、论文中涉及的研究区域或环境类型(绿色)、典范分析的响应变量(蓝色)和解释变量类别(橙色)关键词的共现网络。连线表示关键词的共现关系,节点大小和连线粗细代表了对应文献的数量。典范分析中解释变量的相对重要性分析均来自“rdacca.hp”包。

Fig. 5 Co-occurrence network of subject (red), study area or environment type (green), response variables (blue) and predictor variables (orange) key words according to published papers citing “rdacca.hp” package. Lines represent the co-occurrence relationship of keywords, node size and line width represent the number of related papers. All of the relative importance of predictors in canonical analysis are from “rdacca.hp” package.

性有助于减少土壤养分限制和多功能性的维持。类似地,河湖交错带的植被恢复通过重新分配有机质、氮和生物有效磷组分,缓解水体富营养化增强流域系统的生态功能(Yan *et al.*, 2022b)。仅考虑非生物因子时,有机质组分及氮库变动是影响土壤磷有效性、氮矿化速率以及温室气体排放的重要环境参数(Chen *et al.*, 2021; Kong *et al.*, 2022; Liu *et al.*, 2022a)。由于不可能将这些地球化学变化单独归因于物理化学条件,微生物的活动无疑是联系这些地球化学变化的潜在桥梁。此外,基于HP的分析还展现了特定生物化学标志物在气候或环境建模中的应用价值,例如甘油二烷基甘油四醚表现出与温度、降水和土壤pH的强相关关系(Duan *et al.*, 2022a, 2022b),对于更好地理解生物地球化学循环和气候反馈具有重要意义。

3.3 植物生理

从群落、种群至个体生态学层面,以植物为代表的生理学或表型相关研究也是一大侧重点(图5,图6)。这些典范分析案例中的一个主要部分是了解植物性状变异与气候、环境因子等的相关性。其中,植物叶片作为光合作用的主要器官,在生态系统中起着决定植物生存和生产力的关键作用,其形态结构和营养状态成为研究重点。陆生和水生植物与环境因子的HP均强调了环境过滤在塑造叶片性状方面的重要性(Gao *et al.*, 2021; Ren *et al.*, 2021)。考虑到不同植物类群的叶片性状具有独特的进化和生态价值,叶片性状和环境的关系对于解释和预测叶片营养模式以及识别生态系统营养限制非常重要。淡水的不同理化因子在调节浮游植物、潜水植物以及沉水植物的生态化学计量中的贡献差异巨大(Rao *et*

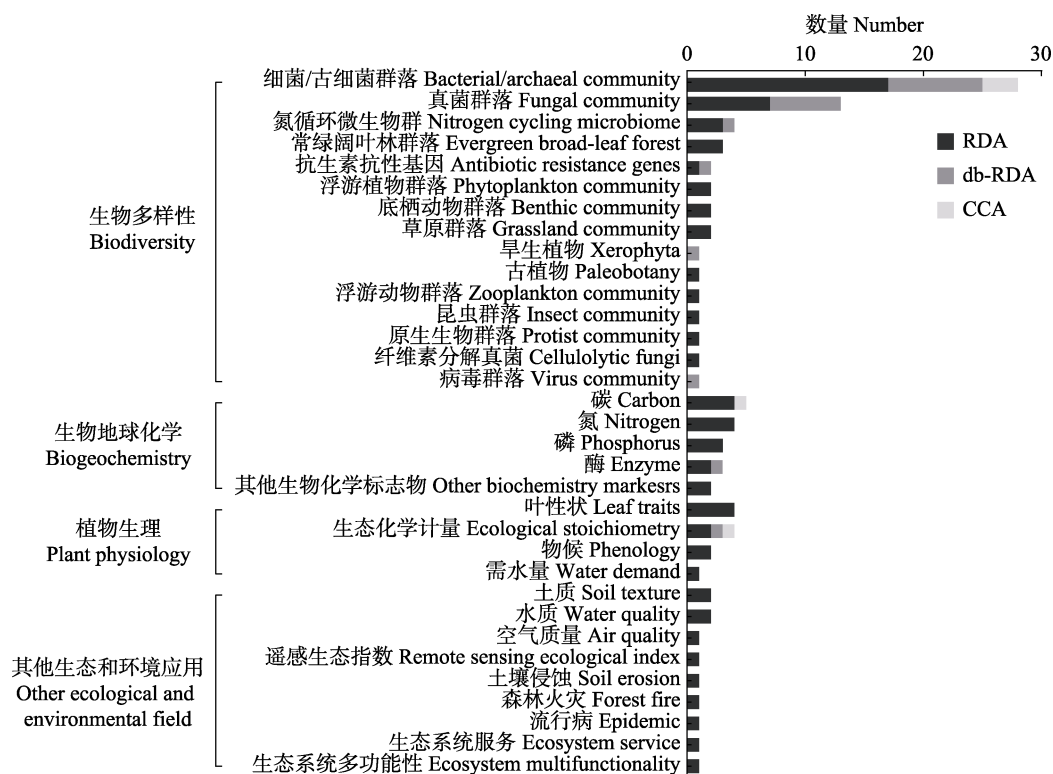


图6 引用“rdacca.hp”包的论文中所涉及典范分析的响应变量类型及数量统计。CCA, 典范对应分析; db-RDA, 基于距离的冗余分析; RDA, 冗余分析。

Fig. 6 Statistics of published papers citing “rdacca.hp” package (type and number of response variables). CCA, canonical correspondence analysis; db-RDA, distance-based redundancy analysis; RDA, redundancy analysis.

al., 2021; Zhang *et al.*, 2021, 2023), 充分体现出3种水生植物的生态位高度分化。另一主要部分是物候学, 量化植物物候及其与气候的相互作用对于应对气候变化和物候变化至关重要(Cleland *et al.*, 2007; Richardson *et al.*, 2013)。HP结果显示, 气候是控制植物形态、物候和生理学变化的主要非生物因子(Zhai *et al.*, 2021; Cheng *et al.*, 2022)。利用功能性状评估沙漠植物开花物候的变异性, 为理解气候变化与植物繁殖策略之间的关系提供了新的视角(Wang *et al.*, 2020)。性状特征还可用于诠释生物多样性。在一个中营养亚热带高原湖泊中, 通过将形态学鉴定和分子生物学的遗传标记分型相结合来揭示浮游植物生态策略, 为制定水华防控措施提供了依据(Hu *et al.*, 2021)。

3.4 其他生态和环境应用

典范分析的其他方面应用还包括对环境指标、自然灾害或生态系统的建模(图5, 图6)。类似于生物地球化学的分析, 环境指标的响应变量亦是一组物理化学参数, 因受人类活动影响强烈而常用做评价环境质量的指标, 如水透明度、空气PM_{2.5}指数等(Tao *et al.*, 2021; Gao *et al.*, 2022; Ke *et al.*, 2022)。了

解这些指标变动的原因有助于制定相应的污染控制措施来降低环境风险。面向自然灾害或生态系统的研究大多上升到景观尺度, 综合调查了区域的自然或人文属性。典范分析对象包括黄土高原的细沟侵蚀、西南森林的火灾时空格局、三峡流域的生态系统服务、黄河流域的土地利用及其与当地的地理地貌、气候、区域环境、生物特征或社会经济驱动因素的联系等(Xiong *et al.*, 2020; Yang *et al.*, 2020; Gou *et al.*, 2021; Cui *et al.*, 2022)。由于因素复杂, 在景观尺度上整合多种生态系统服务仍存在一定挑战。随着学科发展, 景观配置控制的生态系统服务对景观多样性的影响和反馈可能是未来扩大研究的领域。

4 总结和展望

“rdacca.hp”包的HP算法在评估典范分析中共线性的解释变量的相对重要性中表现出优越性, 目前正广泛用于生态学、环境科学和其他相关学科问题的数据分析。本文可进一步加深国内学者对这一方法及其应用的认识。尽管“rdacca.hp”包对输入的解释变量的数量并无限制, 但进一步纳入置换检验评

估HP中单独效应的显著性时,解释变量数量和置换次数的增加使计算过程仍然复杂且耗时(Lai *et al.*, 2022a)。因此,未来有待继续优化算法或开发多线程模式以提高效率。此外,截至当前应用“rdacca.hp”包进行数据分析的公开发表的研究中绝大多数为国人的成果,国外暂且仅两篇引用(图4),未来也有待进一步推广至国际应用。此外,考虑到“rdacca.hp”包无法处理混合效应模型的相对重要性信息提取,基于同一思路的“glmm.hp”包已经发布至CRAN (<https://cran.r-project.org/web/packages/glmm.hp>) 和 GitHub (<https://github.com/laijiangshan/glmm.hp>) (Lai *et al.*, 2022b),进一步完善了这一方法体系。在可预见的未来,这一方法将会成为生态环境统计分析的重要工具。

致谢 感谢南京林业大学水杉英才项目的资助。

参考文献

- Blanchet FG, Legendre P, Borcard D (2008). Forward selection of explanatory variables. *Ecology*, 89, 2623-2632.
- Borcard D, Legendre P (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153, 51-68.
- Borcard D, Legendre P, Drapeau P (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045-1055.
- Burnhan KP, Anderson DR (2004). *Model Selection and Multimodel Inference: a Practical Information-theoretic Approach*. 2nd ed. Springer-Verlag, New York. 10.
- Chen Q, Long CY, Chen JW, Cheng XL (2021). Differential response of soil CO₂, CH₄, and N₂O emissions to edaphic properties and microbial attributes following afforestation in central China. *Global Change Biology*, 27, 5657-5669.
- Cheng M, Wang Y, Zhu JX, Pan Y (2022). Precipitation dominates the relative contributions of climate factors to grasslands spring phenology on the Tibetan Plateau. *Remote Sensing*, 14, 517. DOI: 10.3390/rs14030517.
- Chevan A, Sutherland M (1991). Hierarchical partitioning. *The American Statistician*, 45, 90-96.
- Cleland EE, Chuine I, Menzel A, Mooney HA, Schwartz MD (2007). Shifting plant phenology in response to global change. *Trends in Ecology & Evolution*, 22, 357-365.
- Cox Jr LA (1985). A new measure of attributable risk for public health applications. *Management Science*, 31, 800-813.
- Cui ZQ, Kang HL, Wang WL, Guo WZ, Guo MM, Chen ZX (2022). Vegetation restoration restricts rill development on dump slopes in coalfields. *Science of the Total Environment*, 820, 153203. DOI: 10.1016/j.scitotenv.2022.153203.
- Duan YW, Sun Q, Werne JP, Hou JZ, Yang H, Wang Q, Khormali F, Chen FH (2022a). The impact of precipitation on the distributions of branched tetraethers in alkaline soils. *Organic Geochemistry*, 169, 104410. DOI: 10.1016/j.orggeochem.2022.104410.
- Duan YW, Sun Q, Werne JP, Hou JZ, Yang H, Wang Q, Khormali F, Xia DS, Chu GQ, Chen FH (2022b). General Holocene warming trend in arid Central Asia indicated by soil isoprenoid tetraethers. *Global and Planetary Change*, 215, 103879. DOI: 10.1016/j.gloplacha.2022.103879.
- Gao J, Nunes RF, O'Shea K, Saylor GL, Bu LJ, Kang YG, Duan XD, Dionysiou DD, Luo SL (2022). UV/sodium percarbonate for bisphenol A treatment in water: impact of water quality parameters on the formation of reactive radicals. *Water Research*, 219, 118457. DOI: 10.1016/j.watres.2022.118457.
- Gao YX, Wang L, Hu XQ, Zhang ZL, Liu BG, Zhang XH, Wang GX (2021). Rapid adaptive responses of rosette-type macrophyte *Vallisneria spiralis* juveniles to varying water depths: the role of leaf trait plasticity. *Ecology and Evolution*, 11, 14268-14281.
- Genizi A (1993). Decomposition of R^2 in multiple regression with correlated regressors. *Statistica Sinica*, 3, 407-420.
- Gou MM, Li L, Ouyang S, Wang N, La LM, Liu CF, Xiao WF (2021). Identifying and analyzing ecosystem service bundles and their socioecological drivers in the Three Gorges Reservoir Area. *Journal of Cleaner Production*, 307, 127208. DOI: 10.1016/j.jclepro.2021.127208.
- Graham MH (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84, 2809-2815.
- Grömping U (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61, 139-147.
- Guo X, Feng JJ, Shi Z, Zhou XS, Yuan MT, Tao XY, Hale L, Yuan T, Wang JJ, Qin YJ, Zhou AF, Fu Y, Wu LY, He ZL, van Nostrand JD, *et al.* (2018). Climate warming leads to divergent succession of grassland microbial communities. *Nature Climate Change*, 8, 813-818.
- Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, Gonzalez A, Duffy JE, Gamfeldt L, O'Connor MI (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, 486, 105-108.
- Hu LL, Shan K, Huang LC, Li YR, Zhao L, Zhou QC, Song LR (2021). Environmental factors associated with cyanobacterial assemblages in a mesotrophic subtropical plateau lake: a focus on bloom toxicity. *Science of the Total Environment*, 777, 146052. DOI: 10.1016/j.scitotenv.2021.146052.
- Isbell F, Cowles J, Dee LE, Loreau M, Reich PB, Gonzalez A, Hector A, Schmid B (2018). Quantifying effects of biodiversity on ecosystem functioning across times and

- places. *Ecology Letters*, 21, 763-778.
- Johnson JW, Lebreton JM (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.
- Ke B, Hu WH, Huang DM, Zhang J, Lin XT, Li CH, Jin XJ, Chen J (2022). Three-dimensional building morphology impacts on PM_{2.5} distribution in urban landscape settings in Zhejiang, China. *Science of the Total Environment*, 826, 154094. DOI: 10.1016/j.scitotenv.2022.154094.
- Kong WB, Yao YF, Hou LC, Bao KQ, Zhang LQ, Wei XR (2022). Effects of vegetation presence on soil net N mineralization are independent of landscape position and vegetation type in an eroding watershed. *Agriculture, Ecosystems & Environment*, 325, 107743. DOI: 10.1016/j.agee.2021.107743.
- Kreft H, Jetz W (2007). Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 5925-5930.
- Kruskal W (1987). Relative importance by averaging over orderings. *The American Statistician*, 41, 6-10.
- Lai JS, Zou Y, Zhang JL, Peres-Neto PR (2022a). Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. *Methods in Ecology and Evolution*, 13, 782-788.
- Lai JS, Zou Y, Zhang S, Zhang XG, Mao LF (2022b). glmm.hp: an R package for computing individual effect of predictors in generalized linear mixed models. *Journal of Plant Ecology*, 15, 1302-1307.
- Legendre P, Gallagher ED (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271-280.
- Legendre P, Legendre L (2012). *Numerical Ecology*. 3rd ed. Elsevier, Amsterdam.
- Li S, Yao Q, Liu JJ, Yu ZH, Li YS, Jin J, Liu XB, Wang GH (2022). Liming mitigates the spread of antibiotic resistance genes in an acid black soil. *Science of the Total Environment*, 817, 152971. DOI: 10.1016/j.scitotenv.2022.152971.
- Lindeman RH, Merenda PF, Gold RZ (1980). *Introduction to Bivariate and Multivariate Analysis*. Scott Foresman, Glenview, USA.
- Liu Y, Bing HJ, Wu YH, Zhu H, Tian X, Wang ZG, Chang RY (2022a). Nitrogen addition promotes soil phosphorus availability in the subalpine forest of eastern Tibetan Plateau. *Journal of Soils and Sediments*, 22, 1-11.
- Liu YZ, Shen HH, Ge G, Xing AJ, Tang ZY, Fang JY (2022b). Classification and distribution of evergreen broad-leaved forests in Jiangxi, East China. *Journal of Plant Ecology*, rtac059. DOI: 10.1093/jpe/rtac059.
- Mac Nally R, Walsh CJ (2004). Hierarchical partitioning public-domain software. *Biodiversity and Conservation*, 13, 659-660.
- Mei N, Zhang XZ, Wang XQ, Peng C, Gao HJ, Zhu P, Gu Y (2021). Effects of 40 years applications of inorganic and organic fertilization on soil bacterial community in a maize agroecosystem in northeast China. *European Journal of Agronomy*, 130, 126332. DOI: 10.1016/j.eja.2021.126332.
- Mood AM (1969). Macro-analysis of the American educational system. *Operations Research*, 17, 770-784.
- Newton RG, Spurrell DJ (1967). Examples of the use of elements for clarifying regression analyses. *Journal of the Royal Statistical Society*, 16, 165-172.
- Peres-Neto PR, Legendre P, Dray S, Borcard D (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87, 2614-2625.
- Rao QY, Su HJ, Ruan LW, Deng XW, Wang LT, Rao X, Liu JR, Xia WL, Xu PK, Shen H, Chen J, Xie P (2021). Stoichiometric and physiological mechanisms that link hub traits of submerged macrophytes with ecosystem structure and functioning. *Water Research*, 202, 117392. DOI: 10.1016/j.watres.2021.117392.
- Ren TT, He NP, Liu ZG, Li MX, Zhang JH, Li A, Wei CZ, Lü XT, Han XG (2021). Environmental filtering rather than phylogeny determines plant leaf size in three floristically distinctive plateaus. *Ecological Indicators*, 130, 108049. DOI: 10.1016/j.ecolind.2021.108049.
- Richardson AD, Keenan TF, Migliavacca M, Ryu Y, Sonnentag O, Toomey M (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, 169, 156-173.
- Šmilauer P, Lepš J (2014). *Multivariate Analysis of Ecological Data Using CANOCO 5*. 2nd ed. Cambridge University Press, Cambridge, USA.
- Tao H, Song KS, Liu G, Wen ZD, Wang Q, Du YX, Lyu LL, Du J, Shang YX (2021). Songhua River Basin's improving water quality since 2005 based on Landsat observation of water clarity. *Environmental Research*, 199, 111299. DOI: 10.1016/j.envres.2021.111299.
- van der Plas F (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biological Reviews of the Cambridge Philosophical Society*, 94, 1220-1245.
- Wang F, Kong WD, Ji MK, Zhao K, Chen H, Yue LY, Dong XB (2022a). Grazing greatly reduces the temporal stability of soil cellulolytic fungal community in a steppe on the Tibetan Plateau. *Journal of Environmental Sciences*, 121, 48-57.
- Wang J, Liao LR, Wang GL, Liu HF, Wu Y, Liu GB, Zhang C (2022b). N-induced root exudates mediate the rhizosphere fungal assembly and affect species coexistence. *Science of the Total Environment*, 804, 150148. DOI: 10.1016/j.scitotenv.2021.150148.
- Wang J, Liao LR, Ye ZC, Liu HF, Zhang C, Zhang L, Liu GB,

DOI: 10.17521/cjpe.2022.0314

- Wang GL (2022c). Different bacterial co-occurrence patterns and community assembly between rhizosphere and bulk soils under N addition in the plant-soil system. *Plant and Soil*, 471, 697-713.
- Wang WJ, Wen J, Xiang WQ, Malabrigo Jr PL, Ren MX (2022d). Soil bacterial and fungal communities respond differently to *Bombax ceiba* (Malvaceae) during reproductive stages of rice in a traditional agroforestry system. *Plant and Soil*, 479, 543-558.
- Wang Y, Yang XD, Ali A, Lv GH, Long YX, Wang YY, Ma YG, Xu CC (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, 536. DOI: 10.3389/fpls.2020.00536.
- Xiao X, Xie GX, Yang ZH, He N, Yang D, Liu MD (2021). Variation in abundance, diversity, and composition of nirK and nirS containing denitrifying bacterial communities in a red paddy soil as affected by combined organic-chemical fertilization. *Applied Soil Ecology*, 166, 104001. DOI: 10.1016/j.apsoil.2021.104001.
- Xie L, Chen SF, Feng YY, Li Y, Wang L, He LH, Huang LQ, Wu J, Guo K, Ding H, Fang YM (2022). Mismatch between specific and genetic diversity in an evergreen broadleaf forest in southeast china: a study case of 10.24 ha forest dynamics plot of Huangshan. *Frontiers in Plant Science*, 12, 706006. DOI: 10.3389/fpls.2021.706006.
- Xiong QL, Luo XJ, Liang PH, Xiao Y, Xiao Q, Sun H, Pan KW, Wang LX, Li LJ, Pang XY (2020). Fire from policy, human interventions, or biophysical factors? Temporal-spatial patterns of forest fire in southwestern China. *Forest Ecology and Management*, 474, 118381. DOI: 10.1016/j.foreco.2020.118381.
- Xu HD, Yu MK, Cheng XR (2021). Abundant fungal and rare bacterial taxa jointly reveal soil nutrient cycling and multifunctionality in uneven-aged mixed plantations. *Ecological Indicators*, 129, 107932. DOI: 10.1016/j.ecolind.2021.107932.
- Yan BS, Duan MC, Wang RC, Li JJ, Wei FR, Chen JR, Wang J, Wu Y, Wang GL (2022a). Planted forests intensified soil microbial metabolic nitrogen and phosphorus limitation on the Loess Plateau, China. *Catena*, 211, 105982. DOI: 10.1016/j.catena.2021.105982.
- Yan ZW, Wu L, Lv T, Tong C, Gao ZY, Liu Y, Xing B, Chao CX, Li Y, Wang LG, Liu CH, Yu D (2022b). Response of spatio-temporal changes in sediment phosphorus fractions to vegetation restoration in the degraded river-lake ecotone. *Environmental Pollution*, 308, 119650. DOI: 10.1016/j.envpol.2022.119650.
- Yang SQ, Zhao WW, Pereira P (2020). Determinations of environmental factors on interactive soil properties across different land-use types on the Loess Plateau, China. *Science of the Total Environment*, 738, 140270. DOI: 10.1016/j.scitotenv.2020.140270.
- Zhai DL, Thaler P, Luo YQ, Xu JC (2021). The powdery mildew disease of rubber (*Oidium heveae*) is jointly controlled by the winter temperature and host phenology. *International Journal of Biometeorology*, 65, 1707-1718.
- Zhang B, Tang GL, Luo HL, Yin H, Zhang ZH, Xue J, Huang CB, Lu Y, Shareef M, Gao XP, Zeng FJ (2021). Topsoil nutrients drive leaf carbon and nitrogen concentrations of a desert phreatophyte in habitats with different shallow groundwater depths. *Water*, 13, 3093. DOI: 10.3390/w13213093.
- Zhang Y, Zhang HC, Liu Q, Duan LZ, Zhou QC (2023). Total nitrogen and community turnover determine phosphorus use efficiency of phytoplankton along nutrient gradients in plateau lakes. *Journal of Environmental Sciences*, 124, 699-711.
- Zhao YF, Wang X, Jiang SL, Xiao JJ, Li J, Zhou XH, Liu HY, Hao ZG, Wang KC (2022). Soil development mediates precipitation control on plant productivity and diversity in alpine grasslands. *Geoderma*, 412, 115721. DOI: 10.1016/j.geoderma.2022.115721.

责任编辑: 王少鹏 责任编辑: 乔鲜果