



**UNIVERSIDADE FEDERAL DE GOIÁS**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ECOLOGIA E EVOLUÇÃO**



**Dissertação de Mestrado**

**AVALIAÇÃO DE MÉTODOS PARA ESTIMATIVAS DE  
DISSIMILARIDADE EM GRADIENTES ECOLÓGICOS COM  
ALTA DIVERSIDADE BETA**

**Júlio Cury Hoffmann**

Orientador: Prof. Dr. Adriano Sanches Melo

Goiânia – GO

Abril / 2018

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR  
VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES  
NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico:     Dissertação     Tese

2. Identificação da Tese ou Dissertação:

Nome completo do autor: Júlio Cury Hoffmann

Título do trabalho: AVALIAÇÃO DE MÉTODOS PARA ESTIMATIVAS DE DISSIMILARIDADE EM GRADIENTES ECOLÓGICOS COM ALTA DIVERSIDADE BETA

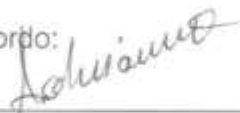
3. Informações de acesso ao documento:

Concorda com a liberação total do documento  SIM     NÃO<sup>1</sup>

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

Júlio Cury Hoffmann  
Assinatura do(a) autor(a)<sup>2</sup>

Ciente e de acordo:

  
Assinatura do(a) orientador(a)<sup>2</sup>

Data: 18 / 05 / 2018

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

<sup>2</sup> A assinatura deve ser escaneada.



**UNIVERSIDADE FEDERAL DE GOIÁS**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ECOLOGIA E EVOLUÇÃO**



**AVALIAÇÃO DE MÉTODOS PARA ESTIMATIVAS DE  
DISSIMILARIDADE EM GRADIENTES ECOLÓGICOS COM  
ALTA DIVERSIDADE BETA**

Júlio Cury Hoffmann

Dissertação apresentada à Universidade  
Federal de Goiás como parte das exigências  
do Programa de Pós-graduação em Ecologia e  
Evolução para obtenção do título de Mestre em  
Ecologia e Evolução.

Orientador: Prof. Dr. Adriano Sanches Melo

Goiânia – GO

Abril / 2018

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Cury Hoffmann, Júlio  
AVALIAÇÃO DE MÉTODOS PARA ESTIMATIVAS DE  
DISSIMILARIDADE EM GRADIENTES ECOLÓGICOS COM ALTA  
DIVERSIDADE BETA [manuscrito] / Júlio Cury Hoffmann. - 2018.  
v, 48 f.: il.

Orientador: Prof. Dr. Adriano Sanches Melo.  
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto  
de Ciências Biológicas (ICB), Programa de Pós-Graduação em Ecologia  
e Evolução, Goiânia, 2018.  
Bibliografia. Anexos.  
Inclui gráfico, tabelas.

1. Esforço amostral. 2. COMPAS. 3. Co-ocorrências. 4. Passo-a  
frente. I. Sanches Melo, Adriano, orient. II. Título.

CDU 574



**SERVIÇO PÚBLICO FEDERAL**  
**UNIVERSIDADE FEDERAL DE GOIÁS - UFG**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS - ICB**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E EVOLUÇÃO - PPGEE**

**ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE Nº 161**

Aos dezoito dias do mês de abril de dois mil e dezoito (18/04/2018), às quatorze horas (14h), no Auditório o ICB V - Campus Samambaia - UFG, reuniram-se os componentes da banca examinadora: **Prof. Dr. Adriano Sanches Melo, ICB-UFG; Prof. Dr. José Alexandre Felizola Diniz Filho, ICB-UFG; Prof. Dr. Rodrigo Augusto Santinelo Pereira, USP**; para, em sessão pública presidida pelo primeiro examinador citado, procederem à avaliação da defesa de dissertação intitulada: "**Avaliação de métodos para estimativas de dissimilaridade em gradientes ecológicos com alta diversidade beta**", em nível de mestrado, área de concentração em Ecologia e Evolução, de autoria de **Júlio Cury Hoffmann**, discente do Programa de Pós-Graduação em Ecologia e Evolução da Universidade Federal de Goiás. A sessão foi aberta pelo presidente, que fez a apresentação formal dos membros da banca. A palavra, a seguir, foi concedida à(o) autor(a) da dissertação que, em cerca de 40 minutos, procedeu à apresentação de seu trabalho. Terminada a apresentação, cada membro da banca arguiu à(o) examinada(o), tendo-se adotado o sistema de diálogo sequencial. Terminada a fase de arguição, procedeu-se à avaliação da dissertação. Tendo-se em vista o que consta na Resolução nº 1127 de dezembro de 2012 do Conselho de Ensino, Pesquisa, Extensão e Cultura (CEPEC), que regulamenta o Programa de Pós-Graduação em Ecologia e Evolução, a dissertação foi APROVADA, considerando-se integralmente cumprido este requisito para fins de obtenção do título de Mestre em Ecologia e Evolução pela Universidade Federal de Goiás. A conclusão do curso dar-se-á quando da entrega da versão definitiva da dissertação na secretaria do programa, com as devidas correções sugeridas pela banca examinadora, no prazo de trinta dias a contar da data da defesa. Cumpridas as formalidades de pauta, às 16 h e 10 min., encerrou-se a sessão de defesa e, para constar,

eu, Suely Ana Ribeiro, secretária executiva da Universidade Federal de Goiás - UFG, lavrei a presente ata que, após lida e aprovada, será assinada pelos membros da banca examinadora em três vias de igual teor.



**Prof. Dr. Adriano Sanches Melo**  
**Presidente da Banca**  
**ICB-UFG**



**Prof. Dr. José Alexandre Felizola Diniz Filho**  
**ICB-UFG**



**Prof. Dr. Rodrigo Augusto Santinelo Pereira**  
**USP**

## **Dedicatória**

Aos meus pais e avôs que ajudaram a me criar e a ser quem sou.

À minha namorada que me atura há muitos anos.

A todos os professores que tanto me ensinaram em todas as etapas onde cheguei.

Ao meu tio Miltinho, (*in memoriam*) pelos grandes momentos que passamos juntos

Ao meu padrinho Rafa e à minha prima “Luíssa”.

Aos meus irmãos e meus amigos quase irmãos, Alexandre e Matheus.

Aos meus queridos Wallace, Cinza, Lu, Ash e Cafezinho

## **Agradecimentos**

Gostaria de agradecer à UFG que propiciou este excelente programa de pós-graduação em Ecologia e Evolução, além da graduação em Ecologia e Análise Ambiental, sem a qual eu não conseguiria a aprovação neste mestrado

À CAPES pelo auxílio financeiro durante todo meu mestrado; Ao professor Adriano Sanches Melo por ter aceitado me orientar e por ter me ajudado e dedicado tanto de seu tempo ao meu aprendizado desde então.

À todos os professores do programa de Ecologia e Evolução que trabalham muito para manter a qualidade desta Pós-graduação, principalmente ao José Alexandre e ao Bini, que se dispuseram a participar da minha qualificação e pelas dicas dadas.

Aos meus colegas de mestrado e de laboratório que estiveram comigo nesta jornada Cléber, Lucas, Herlander, Paula, Vinícius, Flávio, Dani, Luciano, Jean, Angélica, Jesus, Kelly, Anderson, Vini, Victor, Karem, Nayara, Fábio, Daniel, Haunny, Lorena, Marga, Jaques, Alice, Elisa, Lucas, Marco, Leila, Larissa, Renato(s) e especialmente ao André e Danilo, pelas boas conversas que tivemos.

À minha namorada, à minha mãe, à vó Nena e vô Milton e aos meus amigos que me deram apoio fora do ambiente acadêmico e me incentivaram e me ajudaram a atender às exigências durante esses anos do mestrado.

## SUMÁRIO

<b>CAPÍTULO 1: Avaliação de Métodos para Estimativas de Dissimilaridade em Gradientes Ecológicos com Alta Diversidade Beta.....</b>	<b>1</b>
RESUMO .....	1
ABSTRACT .....	2
1. INTRODUÇÃO.....	3
2. MÉTODOS.....	7
2.1 Dados empíricos .....	7
2.2 Dados simulados.....	7
2.3 Análises de dados .....	10
3. RESULTADOS .....	17
3.1 Dados empíricos .....	17
3.2 Dados simulados.....	20
4. DISCUSSÃO .....	31
5. REFERÊNCIAS .....	37
MATERIAL SUPLEMENTAR .....	40
Material Suplementar (I) .....	40
Material Suplementar (II) .....	44

## **Resumo**

Existem vários problemas nas análises de comunidades em Ecologia decorrentes da grande quantidade de zeros na matriz de espécies por locais (dados esparsos), principalmente em situações em que a diversidade beta é alta. Usando quatro estratégias para resolver estes problemas (Beals, Swan, Menor Caminho e Dissimilaridade estendida) retirei aleatoriamente de 1 até 50 % dos indivíduos de matrizes empíricas e simuladas (degradação das matrizes). Realizei ordenações PCoA e nMDS e, com correlações de Procrustes, relacionei a ordenação original em duas dimensões com a ordenação obtida pelas matrizes degradadas. Com dados simulados relacionei a ordenação em duas dimensões também com as coordenadas das amostras no gradiente bidimensional simulado. Além disso, analisei como a robustez à degradação, medida como correlação Procrustes, está relacionada às características da matriz. Diferente do esperado, na comparação das ordenações de matrizes degradadas com a ordenação da matriz original, os dados utilizados sem correções e com índice de dissimilaridade tradicional (Bray-Curtis) tiveram maior ajuste que os quatro métodos avaliados. Em relação às coordenadas, os métodos aplicados tiveram desempenho um pouco melhor do que os dados sem correções quando a diversidade beta foi maior. Em geral, os dados simulados foram mais robustos à degradação que os empíricos e os dados de abundância foram mais robustos que de presença e ausência. Matrizes com menor proporção de zeros foram mais robustas à degradação. Concluo que os métodos de correção distorceram o padrão dos dados originais. Ainda, dados com baixa diversidade beta (poucos zeros na matriz) são robustos à degradação e são suficientes para reconstruir o gradiente original.

## AN EVALUATION OF METHODS TO ESTIMATE DISSIMILARITY IN ECOLOGICAL GRADIENTS WITH HIGH BETA DIVERSITY

### Abstract

There are several problems on the analysis of biological communities with sparse data, resulting from gradients with high beta diversity. I used four strategies to solve this problem (Beals smoothing, Swan, Shortest Path and Extended Dissimilarity). I randomly removed from 1% to 50% of the individuals in empirical and simulated matrices. I then performed PCoA and nMDS ordinations and used Procrustes correlation of the original two dimensional ordination with the ordination obtained using the degraded matrices. For the simulated data set, I also correlated the ordination in two dimensions with the coordinates of the samples in the two-dimensional simulated gradients. Finally, I analyzed how robustness to degradation, quantified as Procrustean correlation, was related to the matrix properties. Different from the expected, in the comparison of the degraded and original ordinations, the uncorrected data with a traditional dissimilarity index (Bray-Curtis) produced higher fit than the four methods evaluated. In relation to the coordinates of the simulated two-dimensional gradients, the evaluated methods were slightly better than the raw data. Overall, the simulated data were more robust to the degradation than the empirical ones and the data of abundance were more robust than matrices of presence and absence. Matrices with small proportion of zeros were more robust to degradation. I conclude that the correction methods evaluated distorted the pattern on the original data. Also, data with low beta diversity (few zeros) are robust to degradation and sufficient to reconstruct the original gradient.

## 1. Introdução

Os dados obtidos em estudos de Ecologia de Comunidades nem sempre são totalmente adequados às técnicas de análises de dados multivariados (McCune 1994, Hirst & Jackson 2007). As matrizes de sítios por espécies incluem grande número de zeros e presença de ruídos. Isso pode ocorrer devido ao esforço amostral não ser suficiente para representar a comunidade de forma acurada. Assim, podem existir espécies não registradas, mas que estão na comunidade (De Cáceres & Legendre 2008). Por outro lado, espécies coletadas podem pertencer a outros habitats, incluindo assim um ruído no entendimento das relações entre as comunidades. Os métodos de análise de semelhança entre as comunidades presumem que as informações fornecidas com os dados de composição e abundância de espécies representam suficientemente as comunidades, o que nem sempre é verdadeiro. Mesmo quando as comunidades são bem amostradas, as matrizes podem incluir muitas espécies, mas poucas em comum. Isto acontece quando há alta diversidade beta e também impossibilita a análise adequada das comunidades (De'ath 1999, Smith 2017).

Métodos multivariados são amplamente usados para mostrar as relações entre objetos a partir de variáveis descritoras (Clarke & Ainsworth 1993). Em Ecologia de Comunidades os objetos geralmente são amostras de comunidades e os descritores são espécies. Os métodos de ordenação multivariada são muito utilizados em Ecologia para descrever padrões de distribuição de espécies. Por exemplo, um gradiente ambiental (como temperatura) pode gerar um gradiente de semelhanças entre comunidades. Ordenações multivariadas servem para posicionar as comunidades conforme as suas semelhanças para que sejam representadas em poucas dimensões (geralmente 2 ou 3). Deste modo, comunidades com composição de espécies e abundâncias relativas parecidas serão posicionadas próximas no espaço da ordenação. O(s) eixo(s) pode(m)

ser comparado(s) com o(s) gradiente(s) estudado(s) (quente-frio) e assim investigar como as comunidades são afetadas pelas características ambientais (Hirst & Jackson 2007, Smith 2017). Outra classe de análise multivariada muito utilizada em Ecologia é classificação ou agrupamento. O objetivo da análise é formar subgrupos a partir de um conjunto de dados, com base nas suas dissimilaridades. Assim, um subgrupo formado com a análise de classificação indica que os seus objetos (ou descritores) são suficientemente mais similares entre si do que com objetos classificados em outros subgrupos. A formação dos subgrupos seguem determinadas regras de aglomeração (classificações aglomerativas) ou divisão (classificações divisivas). Análises de ordenação e classificação começam com o cálculo de uma matriz de dissimilaridade. A escolha do coeficiente de dissimilaridade e das transformações aplicadas são muito importantes para que seja possível revelar o padrão que existe nas relações entre as comunidades (Austin & Greig-Smith 1968, Faith et al. 1987).

Entre as tentativas de resolver os problemas na análise de matrizes de comunidades com grande proporção de zeros estão as transformações de Beals e Swan. A transformação de Beals converte os valores de presença e ausência em probabilidades e, desta forma, reduz a quantidade de zeros na matriz. O índice de Beals é calculado com a seguinte equação:  $b_{ij} = (\sum_{k=1}^S CO_{ik} * I_k / N_k) / S_j$ , onde  $b_{ij}$  é o valor de Beals da espécie  $i$  no local  $j$ ;  $S$  é a riqueza de espécies, sendo que  $S_j$  é a quantidade de espécies no local  $j$ ;  $CO_{ik}$  quantidade de co-ocorrências da espécie  $i$  com a espécie  $k$  na matriz;  $I_k$  valor de incidência da espécie  $k$ , ou seja, se ela está ausente (0) ou presente (1) no local;  $N_k$  quantidade de ocorrências da espécie  $k$  no conjunto de dados. Assim, este valor é a média da probabilidade condicional da espécie em relação às co-ocorrências com as espécies amostradas no local (McCune 1994, Münzbergová & Herben 2004). Existe ainda a opção de excluir a ocorrência da própria espécie no cálculo do índice ( $i \neq k$ );

desta forma o fato da espécie estar presente no local não afeta o valor do seu índice. Esta fórmula pode ser ampliada para dados de abundância de três formas: a) calculando a probabilidade condicional com a abundância ao invés de simplesmente a co-ocorrência (tipo 1), b) utilizando a abundância para dar peso às probabilidades condicionais calculadas com dados binários (tipo 2) e c) com as duas formas ao mesmo tempo (tipo 3) (De Cáceres & Legendre 2008). Além de reduzir a quantidade de zeros na matriz, o índice de Beals também diminui a importância de espécies não relacionadas ao conjunto de dados. Por exemplo, o valor transformado para uma espécie ausente pode ser maior do que para uma espécie que está presente no local, mas que não faz parte da comunidade, ou seja, não tem co-ocorrências com outras espécies da matriz. A transformação de Swan consiste em: a) adicionar 1 e subtrair o menor valor diferente de zero ( $X_{\min} > 0$ ) em todas as observações onde a espécie está presente e b) substituir os zeros pelos valores de Beals. Os dois passos são então repetidos até que não seja possível substituir mais nenhum valor (Oksanen et al. 2017).

Outra estratégia utilizada para analisar comunidades esparsas, nas quais as espécies ocorrem em poucas unidades amostrais, é denominada de passo-a-frente (*step-across*). Ela envolve a modificação direta da matriz de dissimilaridade calculada. São úteis em matrizes com grande quantidade de amostras que não possuem espécies compartilhadas. Existem dois métodos passo-a-frente: menor caminho (*shortest-path*) e dissimilaridade estendida (*extended dissimilarity*). O método de menor caminho se inicia selecionando todos os valores de dissimilaridade acima de um determinado limite (escolhido pelo usuário) ou todas as dissimilaridades iguais a 1. Estes valores são removidos da matriz de dissimilaridade. Em seguida, selecionam-se todas as dissimilaridades que envolvam o par em questão. Soma-se cada uma destas dissimilaridades entre um local intermediário a cada objeto do par em questão que

possua espécies em comum com ambos. A nova estimativa da dissimilaridade é obtida pela seleção do menor somatório entre eles (mínimo somatório das dissimilaridades). O somatório envolvendo estas duas dissimilaridades substituirá a dissimilaridade do par de amostras. O método de dissimilaridade estendida é uma ampliação do método de menor caminho. Após obter as dissimilaridades com o método do menor caminho sucedem outras etapas para substituir os valores de dissimilaridade que ainda não foram estimados. Desta forma, caso ainda exista alguma dissimilaridade acima do limite estabelecido e que não foi substituída, será feita uma nova tentativa utilizando também as dissimilaridades obtidas no passo anterior e assim sucessivamente (De'ath 1999).

Comunidades simuladas são bastante usadas para testar métodos de análise de gradientes, pois com elas é possível saber a priori o padrão que existe entre as comunidades e que se deseja resgatar com a análise (Minchin 1987a, McCune 1994, Hirst & Jackson 2007, De Cáceres & Legendre 2008, Smith 2017), porém não existe consenso quanto à adequabilidade dos modelos aplicados (Austin 2013). Neste trabalho, utilizo dados simulados de comunidades ao longo de gradientes e dados empíricos para avaliar o desempenho em capturar o padrão original dos gradientes em uma ordenação multivariada, quando são aplicadas as transformações das matrizes de comunidades pelos métodos de Beals e Swan e modificações das matrizes de dissimilaridades pelos métodos passo-a-frente. O objetivo é avaliar o desempenho destes métodos quando se reduz o esforço amostral (degradação da matriz original), aumentando a quantidade de zeros e diminuindo a quantidade de espécies compartilhadas entre amostras, situações nas quais as transformações apresentadas são potencialmente necessárias. Isto foi feito a partir de retiradas aleatórias de espécies ou indivíduos em conjuntos de dados simulados e empíricos. Eu então apliquei os métodos citados (Beals, Swan, menor caminho e dissimilaridades estendidas) antes de realizar a análise de ordenação. Comparei o ajuste

dos diferentes métodos para definir quais são melhores e em quais condições eles são úteis.

## **2. Métodos**

### *2.1. Dados empíricos*

Obtive dados de abundâncias de espécies em comunidades no repositório Dryad (datadryad.org), totalizando 43 matrizes. Os dados incluem matrizes de espécies por sítios de diversos grupos taxonômicos obtidos em diferentes estudos (descrição em Material Suplementar I). Também obtive dados de presença e ausência de 291 matrizes reunidas por Atmar & Patterson (1995), onde informações adicionais podem ser encontradas.

### *2.2. Dados simulados*

Obtive matrizes de comunidades a partir da função resposta de espécies a gradientes simulados com base no modelo COMPAS de Minchin (1987b). Grande parte das espécies se distribuem em um gradiente conforme a curva gaussiana (“forma de sino”) ou outras distribuições unimodais relacionadas (Austin 1987, Oksanen & Minchin 2002). Esta observação indica que comunidades geradas com estas curvas representam bem as comunidades reais. Este modelo de simulação começa representando as curvas de abundância de espécies em um ou mais gradiente(s), sendo elas unimodais e semelhantes à curva gaussiana, com variações na simetria e curtose. As matrizes simuladas foram obtidas a partir de amostragens de posições em dois gradientes perpendiculares (locais nos gradientes onde se deseja amostrar as espécies na

simulação). Ou seja, valores de abundâncias (eixo y) de espécies presentes nas coordenadas de dois eixos correspondentes aos gradientes (x1 e x2). Em situações reais, os valores de abundância não seguem tendências lineares ou modais perfeitas. Portanto, os valores de abundância foram substituídos por valores aleatórios de uma distribuição de Poisson com média e variância igual ao valor a ser substituído. Utilizei a função “compas()” do pacote CommEcol (Melo 2016) para obter as matrizes simuladas. A função possui os seguintes argumentos que devem ser fornecidos pelo usuário:

- a) Riqueza potencial (**S**);
- b) Quantidade de gradientes (**dims**);
- c) Abundância (**am**);
- d) Diversidade beta (**beta**);
- e) Localização da amostra no gradiente (**coords**);
- f) Ruído qualitativo (**n.quali**);
- g) Espécies marginais (**add1**).

Os gradientes simulados estão em unidades arbitrárias que variam de 0 a 100, enquanto a moda da curva de respostas de cada espécie é escolhida aleatoriamente entre -50 a 150. Assim, algumas espécies podem ter ocorrência nas extremidades do gradiente e não terem sido amostradas. Como consequência, a riqueza total na simulação pode ser menor que o valor estabelecido pelo parâmetro **S**. O argumento **am** está em escala logarítmica e serve para determinar a abundância no nível ótimo do gradiente para a espécie (moda da curva que representa a abundância). O valor de abundância usado é obtido aleatoriamente de uma distribuição log-normal com média **am** e desvio padrão 1. O uso desta distribuição segue o padrão de comunidades reais, com muitas espécies raras e poucas espécies abundantes.

O argumento **beta** determina a “extensão de ocorrência” (*range*) das espécies ao longo do gradiente e, portanto, é inversamente proporcional à “substituição” (*turnover*) das espécies. Ele é expresso em unidade R, onde  $R = 100/r_{\text{médio}}$ . O valor r representa a “extensão de ocorrência” da espécie e é obtido de uma distribuição normal com média igual a  $100 \cdot R$  e desvio padrão igual a  $0,3 \cdot 100 \cdot R$ . Portanto, quanto maior o valor do parâmetro beta no modelo COMPAS, maior será a extensão de ocorrência média das espécies e menor será a diversidade beta.

O parâmetro **add1** adiciona espécies que não fazem parte do conjunto regional, mas que eventualmente têm uma única ocorrência em uma comunidade, na proporção de **add1**. O **n.quali** é o ruído qualitativo que serve para substituir a abundância de uma espécie em um local por zero, na proporção de **n.quali**. Assim, quanto maior o valor de **add1** mais espécies que não são adequadas ao local devem ocorrer e quanto maior o **n.quali** mais espécies que deveriam estar em uma comunidade não estarão.

Utilizei os seguintes parâmetros: riqueza máxima (**S**) = 90 até 250 (ver abaixo); gradientes ortogonais (**dims**) = 2; abundância (**am**) = 2 (para os dois gradientes); substituição (**beta**) = variável (ver abaixo); coordenadas de cada gradiente (**coords**) = 15, 25, ..., até 85, para as duas dimensões, de modo que se forme uma grade de 8 x 8 e totalizando 64 unidades amostrais (Fig. 1); proporção de espécies marginais (**add1**) = 0,01 e proporção de ruído qualitativo (**n.quali**) = 0,05.

Escolhi 4 valores para o parâmetro **beta**, de 0,4 até 1,0 em passos de 0,2. Criei 30 matrizes para cada nível de diversidade beta. Assim, obtive 120 matrizes (30 matrizes para cada um dos 4 níveis de diversidade beta). Desta forma, os conjuntos de matrizes simuladas serão mencionados de acordo com o aumento na diversidade beta do nível 1 ao 4. Aumentei o valor da riqueza potencial (**S**) quando diminuí o valor do

parâmetro **beta**, de modo que a riqueza total das matrizes nos 4 níveis de diversidade beta não seja muito diferente. Assim, pude avaliar o efeito da diversidade beta, visto que os métodos avaliados supostamente são melhores em situações de alta diversidade beta, mas controlando o efeito da diversidade gama.

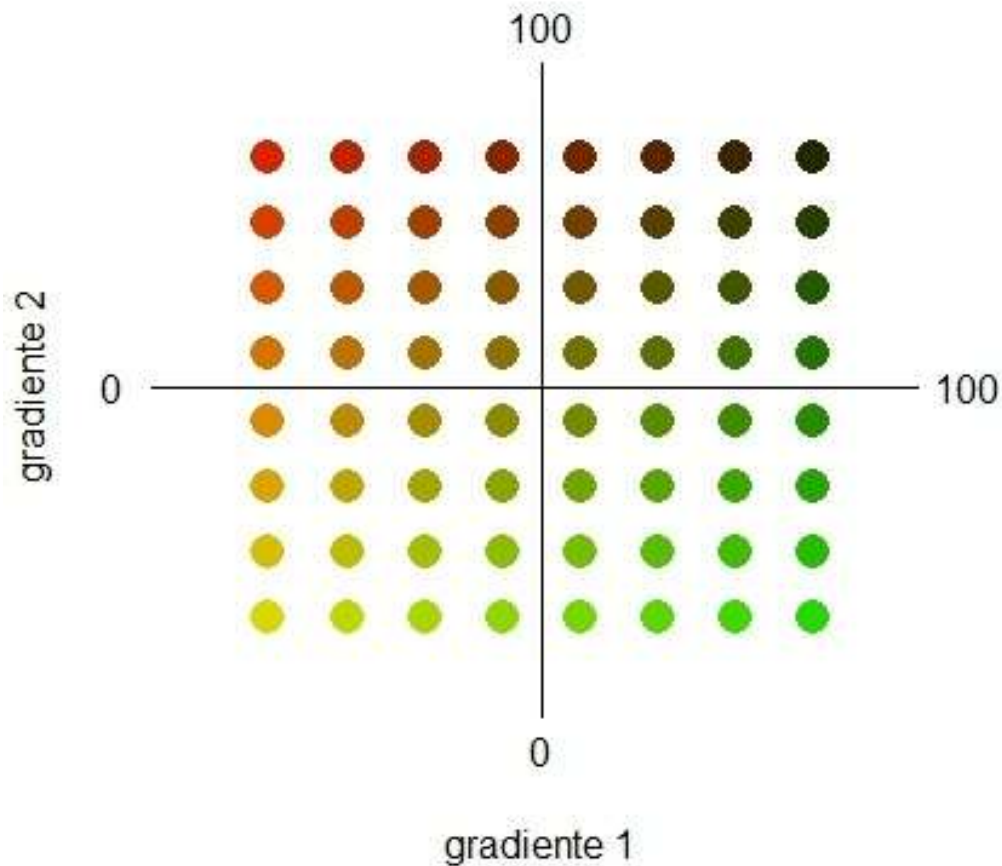


Fig. 1: Disposição das comunidades nos gradientes da simulação. Cada círculo representa uma amostra obtida (comunidade local). Estão apresentadas as combinações dos dois fatores (gradientes). As comunidades estão igualmente distribuídas com valores altos para os dois fatores e baixos para os dois fatores, bem como a combinação de valores altos e baixos em cada um dos gradientes.

### 2.3. Análises de dados

Realizei a remoção aleatória de indivíduos utilizando um código que eu criei dentro do programa R (R Core Team 2016). Em dados de presença e ausência, retirar um indivíduo significa a remoção da incidência da espécie em um local. Assim, removi sucessivamente a proporção de 1% até 50% da soma total das abundâncias ou das incidências (dados qualitativos) da matriz. Para avaliar apenas os métodos, também obtive matrizes de dissimilaridade transformadas sem degradação (degradação = 0%). Para que a quantidade de unidades amostrais não fosse alterada mantive pelo menos um indivíduo (ou incidência) em cada linha. As matrizes com menos de 100 incidências no total foram excluídas da análise. Assim, analisei 158 matrizes das 291 matrizes empíricas de presença e ausência obtidas. Com as matrizes de abundância, realizei também as degradações das matrizes utilizando apenas as incidências, como feito com as matrizes de presença e ausência. Para isso, utilizei as 23 matrizes de abundância com mais de 100 incidências.

Apliquei uma transformação logarítmica nos dados de abundâncias,  $\log_{10}(\text{abundância} + 1)$ . Em seguida, transformei as matrizes com os índices de Beals, incluindo a ocorrência da própria espécie (i.e. “include = TRUE”, na função `beals()` do pacote `vegan`) e Swan antes de calcular a matriz de dissimilaridade de Bray-Curtis (ou Sorensen no caso de dados de presença-ausência). Os valores de abundância (*log*) foram utilizados no índice de Beals para dar peso às probabilidades condicionadas calculadas (tipo 3, i.e. `type = 3` na função `beals()`, disponível no pacote `vegan`). Também obtive as dissimilaridades passo-a-frente (*step-across*) com base nas matrizes de dissimilaridade de Sorensen e Bray-Curtis, calculados com dados de presença-ausência ou *log* da abundância, substituindo os valores de dissimilaridades maiores que 0,95 com os métodos de menor caminho e dissimilaridades estendidas.

Finalmente, para avaliar os métodos, calculei a dissimilaridades de Bray-Curtis ou sua versão binária (Sorensen) sem aplicar qualquer uma das estratégias de correção de dados com alta diversidade beta. Para os dados empíricos, portanto, obtive 5 matrizes de dissimilaridades para cada matriz de dados qualitativos (4 métodos + 1 sem método) e 10 para cada matriz de abundância (as mesmas cinco anteriores para dados de abundância e de presença-ausência). Como os métodos passo-a-frente envolvem substituição de valores ausentes na matriz, em alguns casos não é possível calcular a dissimilaridade para todos os valores e a matriz de dissimilaridade resultante possui valores ausentes. Desta forma, removi da análise todas as matrizes de espécies que produziu valores ausentes de dissimilaridade em qualquer nível de degradação (Fig. 2). Isso ocorreu com 18 das 158 matrizes de presença e ausência restantes. Assim, também removi 3 das 43 matrizes de abundância e quando analisei estas matrizes novamente apenas com as presenças e ausências removi 5 das 23 matrizes com mais de 100 incidências. Portanto, os resultados apresentados com dados empíricos são referentes a 140 matrizes originalmente de presença ou ausência, 40 matrizes de abundância e 18 matrizes originalmente com dados de abundância que foram transformadas em presença e ausência.

Com os dados simulados também apliquei para cada matriz de comunidade as 4 transformações (Beals, Swan, menor caminho e dissimilaridade estendida) além da matriz de dissimilaridade de Bray-Curtis sem modificações. Assim como fiz com dados empíricos de abundância, também analisei as matrizes simuladas com as presenças e ausências. Deste modo refiz a degradação das matrizes apenas com as incidências, apliquei os 4 métodos citados e comparei também com os resultados obtidos utilizando dissimilaridade de Sorensen sem modificações. Excluí as matrizes que possuíam valores ausentes de dissimilaridade com os métodos passo-a-frente e, neste caso, obtive outras

matrizes através de simulações para manter sempre 30 matrizes em cada nível de diversidade beta.

Com estas matrizes de dissimilaridades obtive ordenações multivariadas, por meio de Análise de Coordenadas Principais (PCoA) e Escalonamento Multidimensional não-métrico (nMDS) com até 20 inícios aleatórios, para todas as porcentagens de degradação das matrizes (Fig. 2). O nMDS foi feito com dois eixos. Isto resultou, para cada matriz original de dados empíricos qualitativos, em 510 ordenações (5 matrizes de dissimilaridade \* 51 porcentagens de degradação \* 2 métodos de ordenação). Para cada matriz de abundância, obtive 1020 ordenações (510 \* 2 tipos de dados [presença-ausência e abundância]). Da mesma forma, também obtive, para os dados simulados, 1020 ordenações (5 matrizes de dissimilaridade \* 51 porcentagens de degradação \* 2 métodos de ordenação \* 2 tipos de dados) para cada matriz de comunidade.

Em seguida, comparei as ordenações das matrizes degradadas com a ordenação da matriz original (inteira, sem qualquer estratégia de correção), utilizando os valores de ajuste ou correlação de Procrustes (Fig. 2). Com as matrizes simuladas, também relacionei por meio de Procrustes a estrutura obtida na ordenação em dois eixos principais às coordenadas dos locais nos dois gradientes (Fig. 2). O teste Procrustes foi escolhido pois fornece uma medida da similaridade geral dos resultados, comparando a estrutura geral das matrizes (Jackson 1993). Para sumarizar a robustez do ajuste ao longo da degradação, fiz uma regressão linear com os valores de correlação da análise Procrustes (51 valores de correlação = 1 matriz intacta + 50 degradações) como variável resposta em relação à porcentagem de degradação da matriz (0% até 50%). Assim, obtive o coeficiente de inclinação (b) para cada matriz de dados (Fig. 3). Um bom método produzirá valores altos, ou seja, valores próximos de 0. Um método ruim não será capaz de recuperar o padrão a partir de matrizes degradadas e, portanto, a

correlação Procrustes deverá ser reduzida fortemente conforme se degrada a matriz (valores negativos distantes de 0).

Obtive para cada conjunto de dados (1) se os dados são presença e ausência ou de abundância, (2) quantas unidades amostrais existem no estudo, (3) riqueza de espécies, (4) média de riqueza por unidade amostral, (5) a proporção de espécies com um ou (6) dois indivíduos, (7) proporção de zeros, (8) a média da dissimilaridade (qualquer que seja o tipo de dado na matriz, presença-ausência ou abundância) e (9) a proporção das dissimilaridades maiores que 0,95. Posteriormente, relacionei estas propriedades dos conjuntos de dados ao desempenho dos métodos avaliados.

Eu então relacionei estas inclinações com as propriedades das matrizes através de uma regressão múltipla para cada método utilizado, tanto para dados empíricos como para dados simulados. Para os dados simulados, não utilizei a variável referente ao número de unidades amostrais, uma vez que foi igual em todas as simulações. Antes de realizar a análise de regressão múltipla padronizei as variáveis descrevendo as propriedades das matrizes de dados, de modo que cada variável passou a ter média 0 e desvio padrão 1. Realizei também análises de regressão múltipla entre o valor de ajuste da matriz intacta (0% de degradação), obtidos com os métodos Beals, Swan, menor caminho e dissimilaridades estendidas, e suas variáveis padronizadas (propriedades da matriz empírica). Da mesma forma, com os dados simulados fiz uma regressão múltipla para cada método, utilizando os valores de Procrustes da matriz original (sem degradações) em função das propriedades das matrizes simuladas. Para os dados simulados também fiz as análises de regressão múltipla para os coeficientes de inclinação obtidos em relação às coordenadas ( $x_1$ ,  $x_2$ ) nos gradientes. Todas as análises foram feitas com o pacote vegan (Oksanen et al. 2017) ou rotinas que eu escrevi em R.

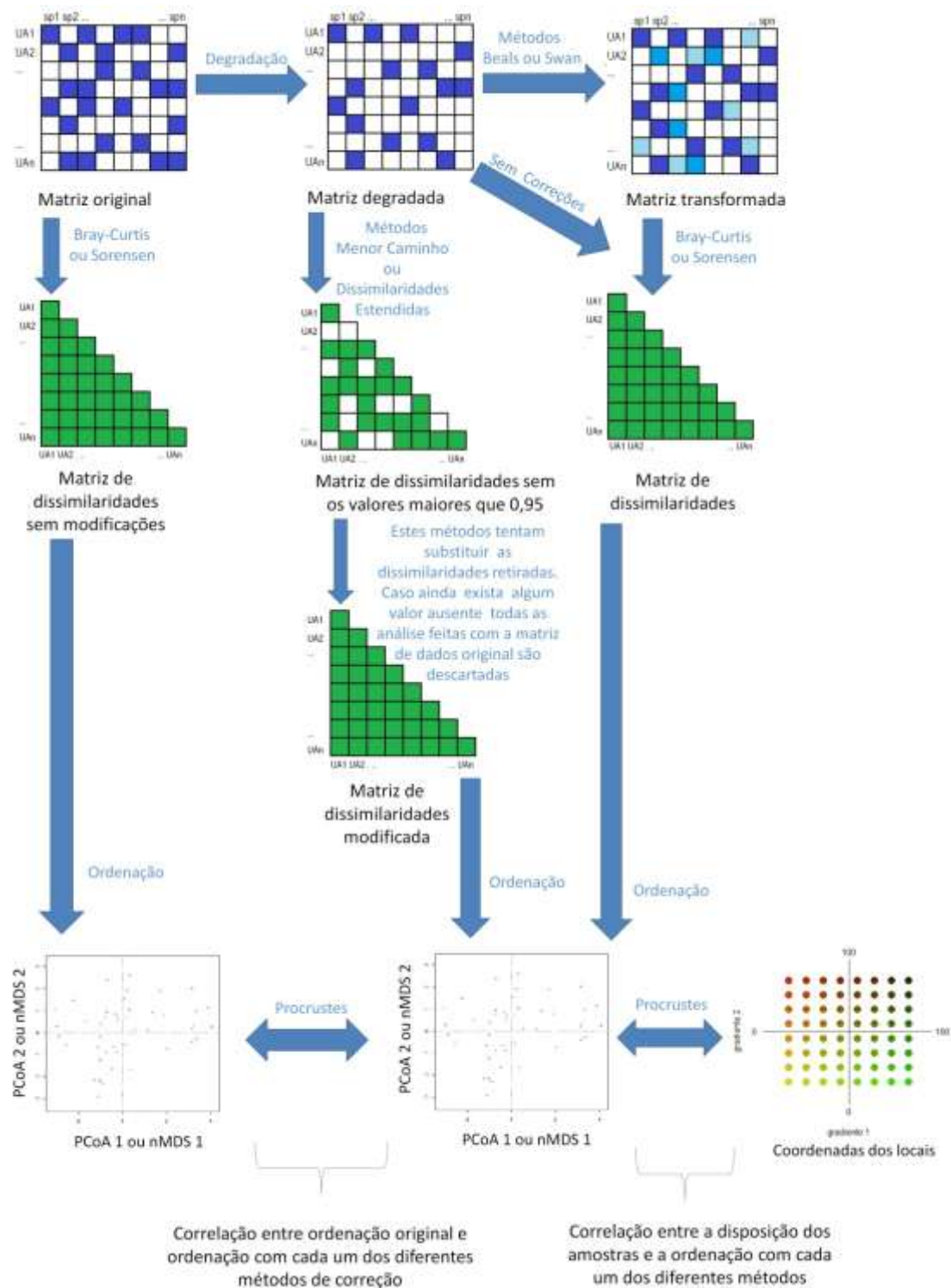


Fig. 2: Representação esquemática dos diferentes métodos aplicados em uma matriz de comunidades degradada (matrizes retangulares com amostras nas linhas e espécies nas colunas), transformações aplicadas em matrizes de dissimilaridades (matrizes triangulares de amostras por amostras) e subsequentes comparações realizadas com Procrustes.

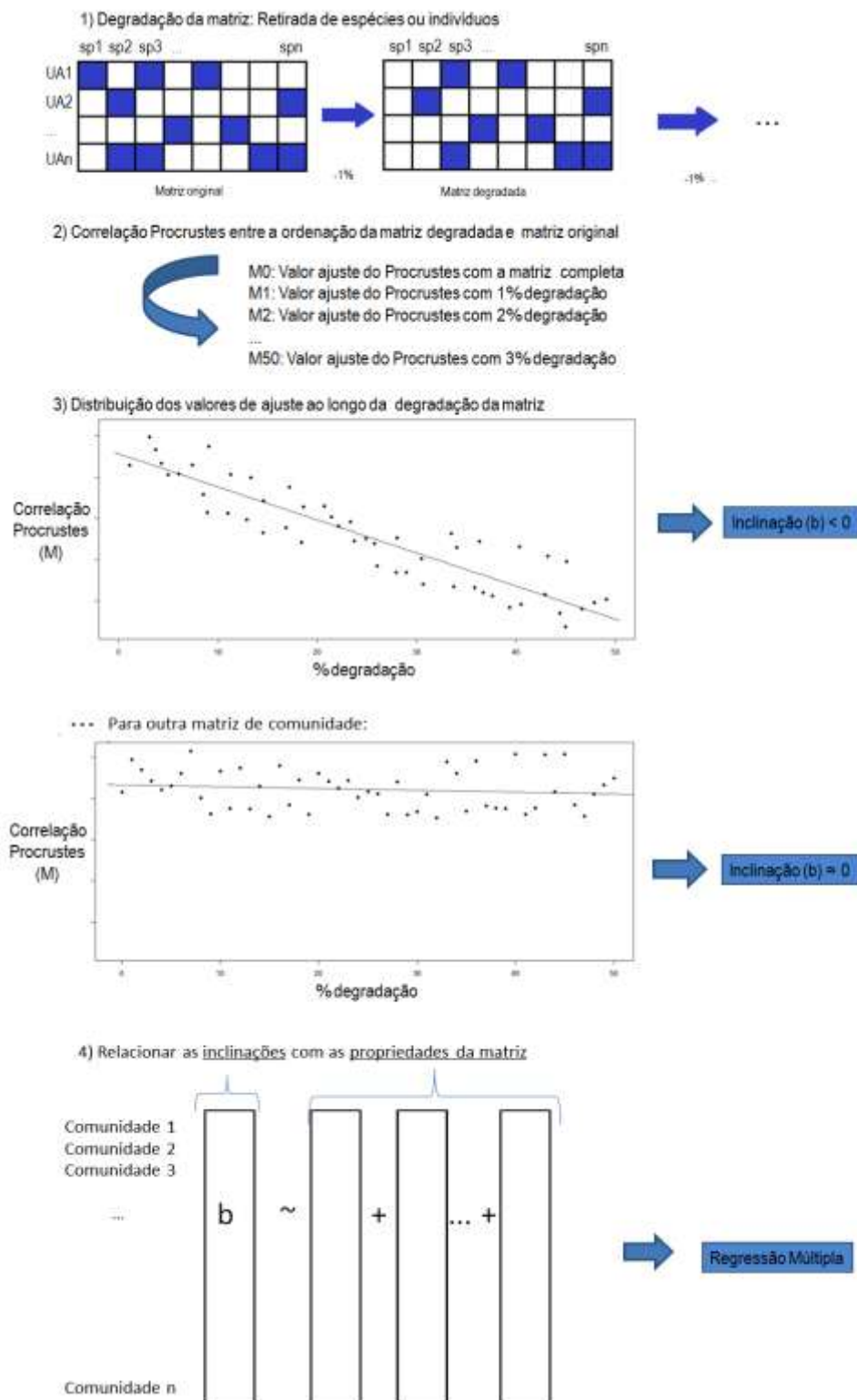


Fig. 3: Esquema mostrando as etapas para análise do desempenho dos métodos de correção para conjuntos de dados esparsos em relação a degradações da matriz de comunidade. Para cada conjunto de dados, com um determinado método, o procedimento consistiu em 1) degradar a matriz original, 2) obter o valor do ajuste de

Procrustes para cada porcentagem de degradação, 3) obter a relação linear entre os ajustes de Procrustes e as porcentagens de degradação e 4) analisar a relação entre a inclinação (b), obtida na etapa anterior, e as propriedades da matriz de comunidade.

### **3. Resultados**

#### *3.1 Dados empíricos*

Apresento a seguir os resultados com as matrizes empíricas de abundância, com as matrizes originalmente de presença e ausência e também com essas matrizes de abundância analisadas utilizando as informações de presença e ausência. Estes resultados são referentes à técnica de ordenação PCoA, pois a correlação de Procrustes média dos diferentes métodos foi maior do que utilizando o método de nMDS.

*Dados empíricos de abundância.* A análise feita com as matrizes de comunidades e de dissimilaridade sem modificações, ou seja, sem nenhuma estratégia de correção, teve a maior qualidade de ajuste (correlação Procrustes) com o padrão original de ordenação (Fig. 4A). Entre as estratégias que apliquei para correção desses dados, os métodos passo-a-frente tiveram desempenhos melhores do que as transformações de Beals e Swan. A transformação de Beals teve correlação consistentemente maior que o método de Swan.

*Dados empíricos de presença e ausência.* Assim como o resultado para dados de abundância, o desempenho utilizando matrizes sem correções foi maior do que com métodos passo-a-frente, Beals e Swan, sucessivamente. O desempenho com dados de presença e ausência diminuiu substancialmente com a degradação das matrizes, ou seja, a perda de qualidade do ajuste com dados de abundância foi menor. Após sucessivas

degradações, as matrizes transformadas conseguem recuperar o padrão original (se aproximam do resultado com a matriz sem transformações). Nos últimos níveis de degradação os métodos tem desempenho semelhantes (Fig. 4B)

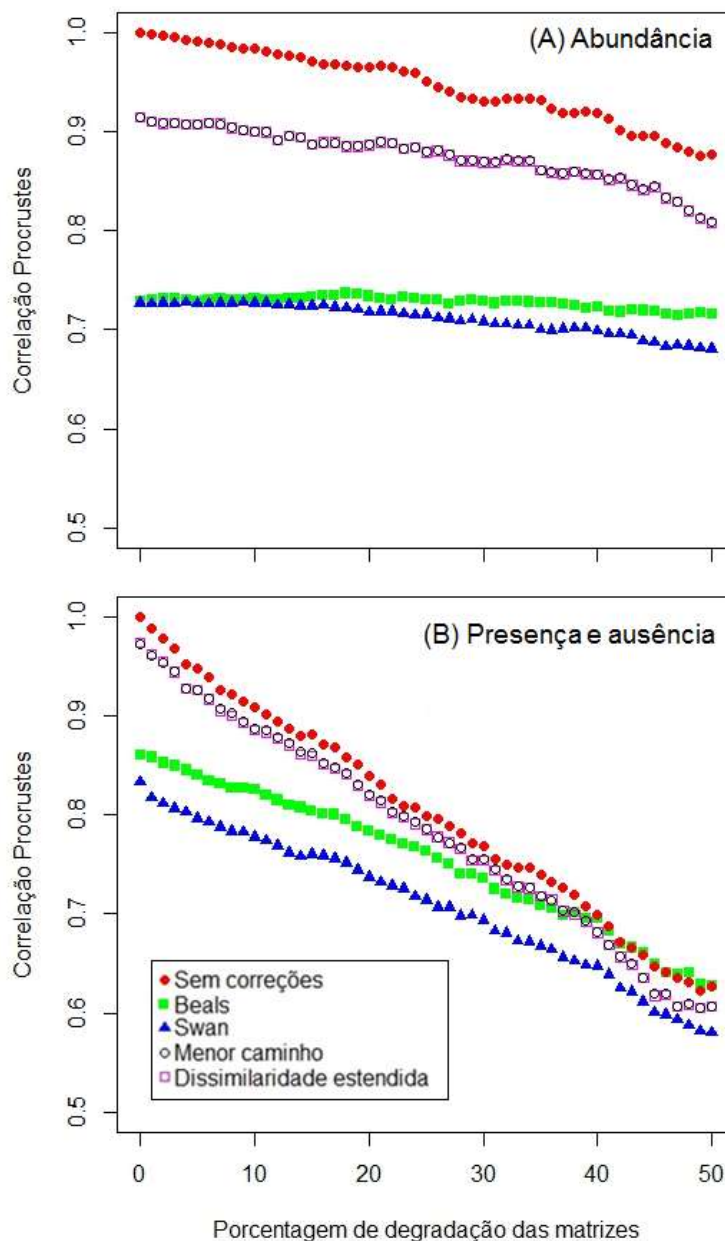


Fig. 4: Comparação das estratégias utilizadas para correção de dados com grande quantidade de valores ausentes (i.e. transformação de Beals e Swan e métodos menor caminho e dissimilaridade estendida) e matrizes degradadas sem aplicar qualquer estratégia de correção em relação à suas respectivas matrizes originais completas. Estão apresentados os valores dos ajustes de Procrustes de cada estratégia em relação à

porcentagem de degradação das matrizes. Os dados de abundância (A) representam as médias dos ajustes de 40 matrizes empíricas e os dados de presença e ausência (B) são compostos de 140 matrizes originalmente de presença e ausência e 17 matrizes transformadas para presença e ausência das 40 matrizes de abundância.

*Efeito das propriedades da matriz em relação à inclinação do ajuste ao longo da degradação.* De modo geral, as matrizes empíricas foram mais robustas à degradação quando elas possuíam dados de abundância (tipo de dados), maiores proporções de espécies com uma ou duas ocorrências, maior proporção de zeros na matriz original e maior dissimilaridade média (Tabela 1). *Efeito das propriedades da matriz em relação à correlação Procrustes das matrizes sem degradação.* As matrizes com maior ajuste Procrustes foram aquelas com dados de abundância (com método de Swan), com mais unidades amostrais (Beals, Swan e Dissimilaridades estendidas), maior proporção de espécies com uma ocorrência (Swan, Menor caminho e Dissimilaridades estendidas), menores proporções valores zeros (Beals, Swan e Menor Caminho) e menor dissimilaridade média e menor proporção de dissimilaridades maiores que 0,95 (Tabela 1).

Tabela 1: Resultados da regressão relacionando a robustez das matrizes à degradação com as suas propriedades. Apresento também a análise de regressão entre o ajuste Procrustes da matriz sem degradações e as propriedades da matrizes. Estão representados apenas os resultados significativamente relacionados e se a relação é positivas (+) ou negativa (-).

---

Robustez à degradação

Tratamento dos dados	Tipo de dados	Riqueza total	Proporção de <i>singletons</i>	Proporção de <i>doubletons</i>	Proporção de zeros na matriz	Dissimilaridade média
Brutos	+	+	+		-	+
Beals	+		+		-	+
Swan	+		+		-	+
Menor caminho	+			+	-	+
Diss. estendidas	+			+	-	+

---

Correlação matriz
sem degradação

Tratamento dos dados	Tipo de dados	Número de Locais (U.A.)	Proporção de <i>singletons</i>	Proporção de zeros na matriz	Dissimilaridade média	Proporção de dissim. > 0,95
Beals		+		-		
Swan	+	+	+	-	-	-
Menor caminho			+	-		
Diss. estendidas		+	+			

---

### 3.2 Dados simulados

Utilizei 30 matrizes para cada nível de diversidade beta (Tabela 2). Os resultados apresentados são referentes à técnica de ordenação nMDS, visto que ela teve o ajuste Procrustes médio maior do que com as ordenações PCoA.

*Dados simulados de abundância.* Assim como o observado com os dados empíricos, os dados simulados sem correções tiveram maior qualidade de ajuste Procrustes que as análises com os 4 métodos de correção de dados esparsos. Com a diversidade beta baixa os métodos tiveram desempenho bastante semelhantes (Fig 5A-B). Quando a diversidade beta foi maior o método de Swan teve desempenho menor que os demais, que apresentaram desempenho semelhantes (Fig. 5C-D).

Tabela 2: Descrição das matrizes simuladas de acordo com os valores de diversidade gama média, mínima e máxima em cada um dos níveis de diversidade beta, assim como a riqueza média por local (alfa média) das matrizes.

Diversidade Beta	Diversidade Gama	Diversidade Alfa (média por local)
Nível 1	75,6 (68 – 85)	15,9
Nível 2	73,0 (62 – 83)	11,7
Nível 3	79,9 (65 – 97)	9,7
Nível 4	75,7 (59 – 88)	5,6

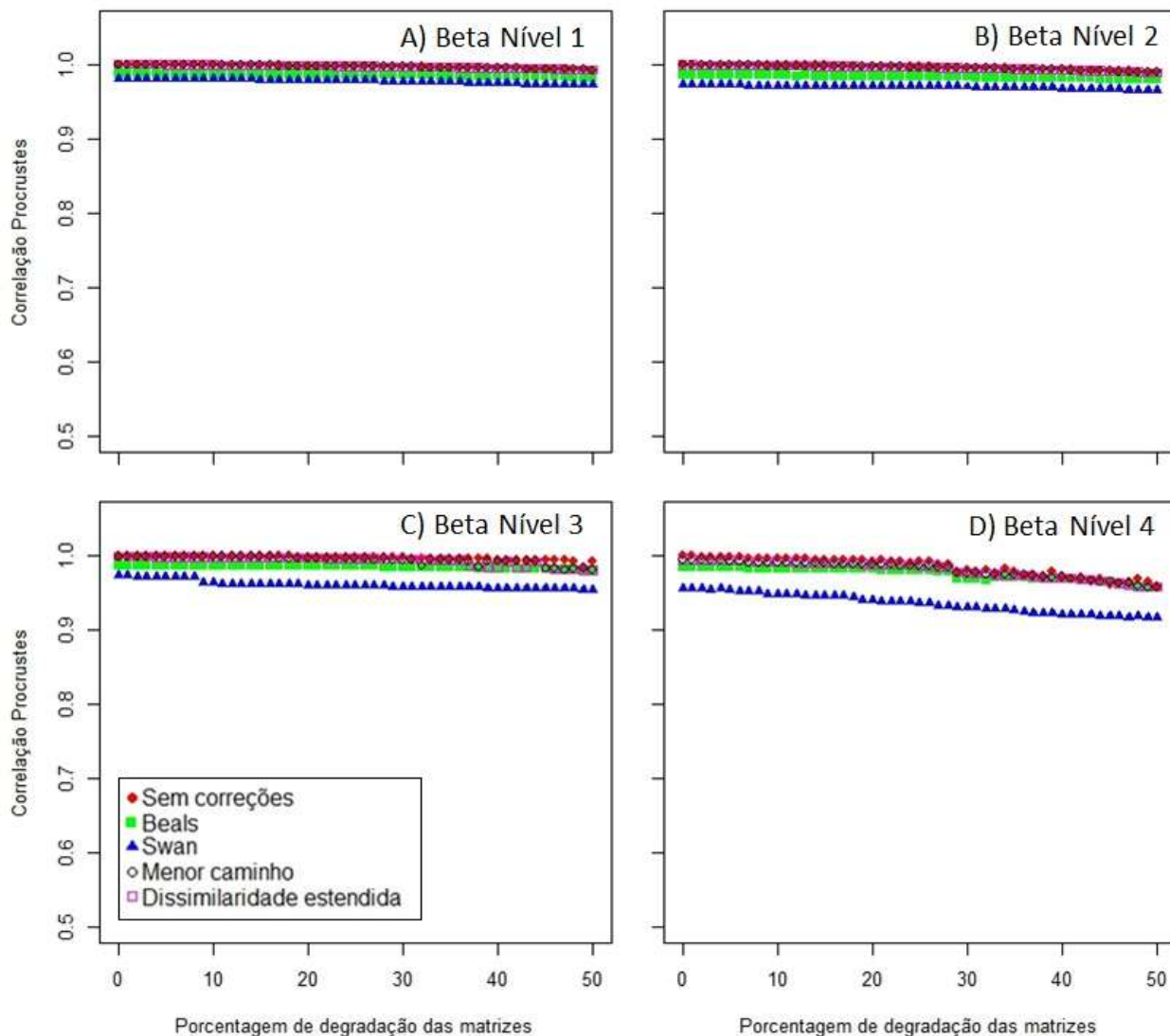


Fig. 5: Média do ajuste Procrustes das matrizes degradadas simuladas em relação à ordenação da matriz original. Em cada diagrama estão as curvas das 5 formas analisadas para obter a dissimilaridade entre locais, cujos conjuntos são compostos por 30 matrizes em cada nível de diversidade beta (determinado pelo parâmetro beta do modelo), sendo: nível 1 (A) com as matrizes cuja diversidade beta é menor até o nível 4 (D) com a maior diversidade beta.

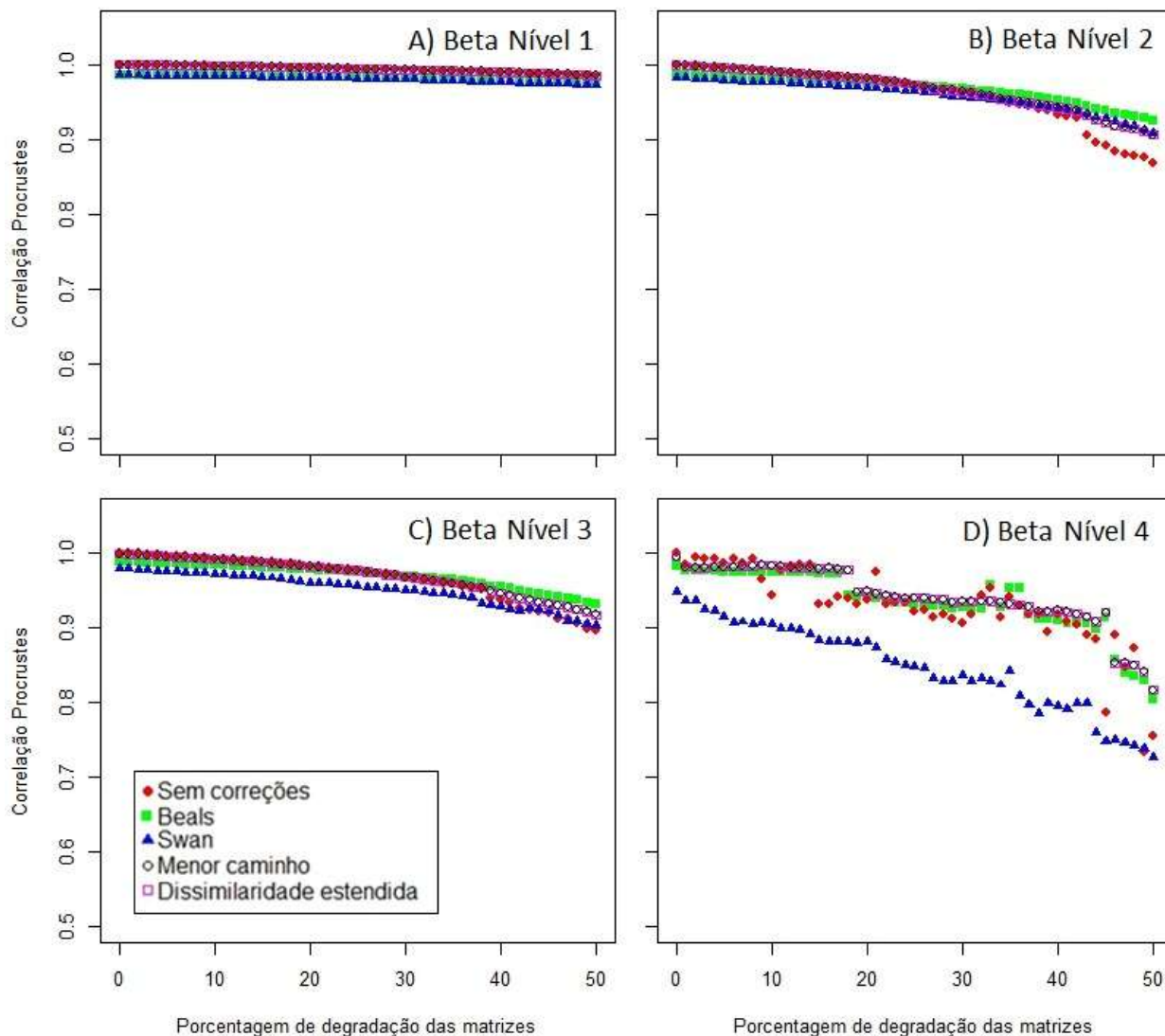


Fig. 6: Média dos valores de ajuste do teste Procrustes com as matrizes simuladas também em relação à ordenação original, com a degradação das matrizes com base na presença ou ausência de espécies. Como as matrizes utilizadas são as mesmas da figura anterior, os níveis de diversidade beta continuam os mesmos. Porém no nível 2 (B) uma matriz foi retirada, enquanto no nível 4 (D) restaram apenas 3 matrizes sem valores ausentes na dissimilaridades. Desta forma estes níveis representam um subconjunto das suas respectivas matrizes de abundância.

*Dados simulados de presença-ausência.* Com baixa diversidade beta os métodos são bastante semelhantes (Fig. 6A-B). As transformações de Beals e Swan foram

robustas à degradação das matrizes. Quando a diversidade beta aumentou, os métodos através passos e sem correções tiveram ajuste maior nas primeiras degradações, sendo inclusive superados quando realizadas sucessivas degradações (Fig. 6C-D). Assim como com dados empíricos, as matrizes de presença e ausência têm alta relação com as matrizes originais quando o nível de degradação foi baixo. Ainda as matrizes de presença e ausência foram mais sensíveis à degradação (Fig. 6) do que aquelas com abundância (Fig. 6).

*Efeito das propriedades da matriz em relação à inclinação do ajuste ao longo da degradação.* As matrizes mais robustas a degradação foram as que possuíam dados de abundância; com maior riqueza total; menor riqueza média por unidade amostral (exceto com o método de Swan); menor proporção de *singletons* ou *doubletons*; menor proporção de valores zero; menor dissimilaridade média (exceto dados sem modificações) e maior proporção de dissimilaridades maiores que 0,95 (Tabela 3).

*Efeito das propriedades da matriz em relação à correlação Procrustes das matrizes sem degradação.* Apenas com o método de Menor Caminho e Dissimilaridades estendidas as matrizes com maior correlação Procrustes foram aquelas com maior dissimilaridade média e menor proporção de dissimilaridades maiores que 0,95 (Tabela 3).

Tabela 3: Resultados da regressão entre a robustez das matrizes simuladas à degradação e as suas propriedades. E da análise de regressão entre o ajuste Procrustes das matrizes sem degradações e as propriedades da matrizes. Estão representados apenas os resultados significativos e se a relação é positiva (+) ou negativa (-).

Robustez à degradação

Tratamento dos dados	Tipo de dados	Riqueza total	Riqueza média por U.A.	Proporção de <i>singletons</i>	Proporção de <i>doubletons</i>	Proporção de zeros na matriz	Dissimilaridade média	Proporção de diss. > 0,95
Brutos	+	+	-	-	-	-		+
Beals	+	+	-	-	-	-	-	+
Swan	+	+			-	-	-	+
Menor caminho	+	+	-		-	-	-	+
Diss. Estendidas	+	+	-		-	-	-	+

Correlaçãomatrizsem degradação

Tratamento dos dados	Dissimilaridade média	Proporção de dissimilaridades > 0,95
Beals		
Swan		
Menor caminho	+	-
Diss. Estendidas	+	-

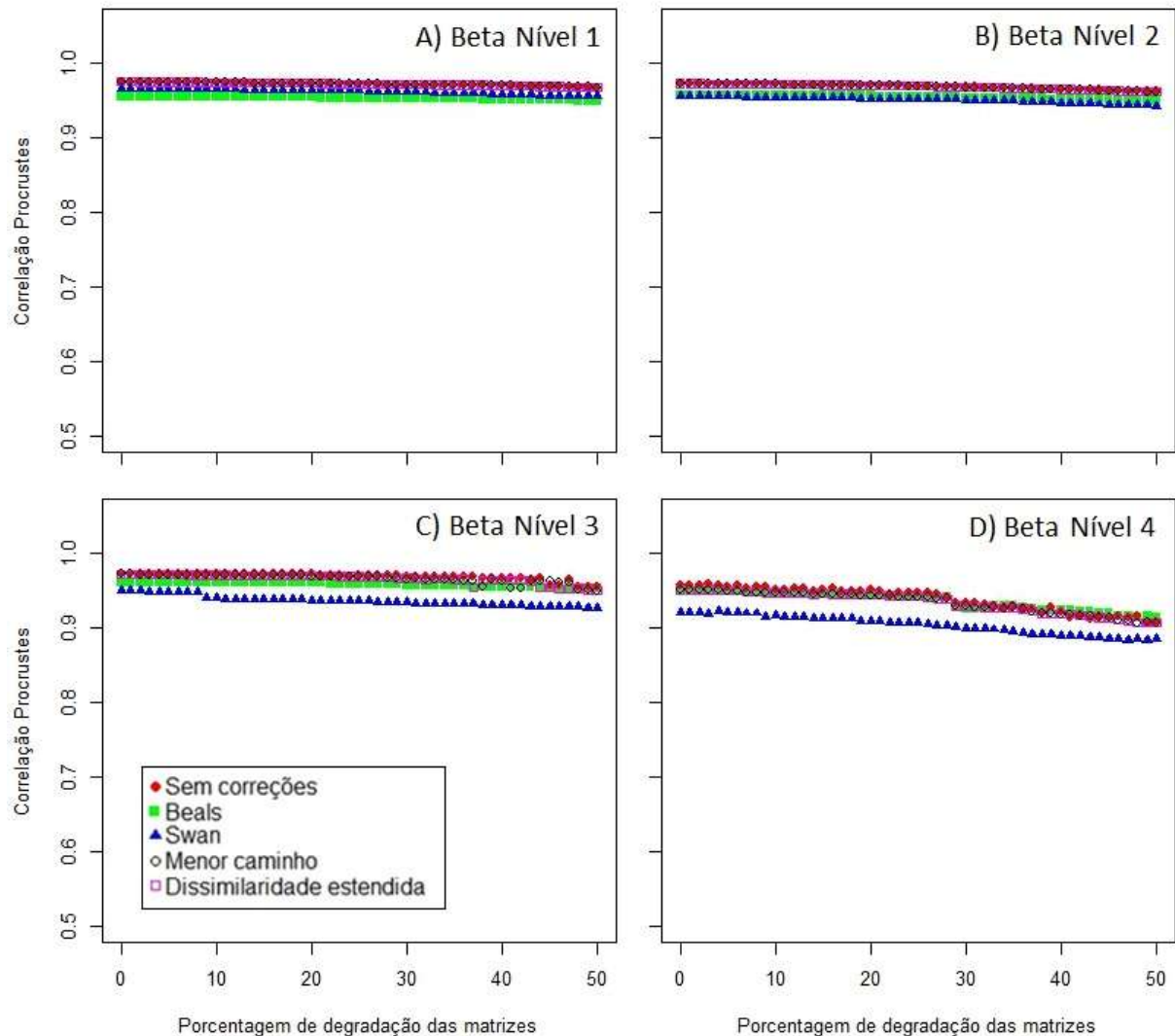


Fig. 7: Média do ajuste da correlação de Procrustes em relação às coordenadas das unidades amostrais da simulação com 5 formas de medir dissimilaridade (Bray-Curtis, Beals, Swan, Menor Caminho e dissimilaridades estendidas) usando dados de abundância. O conjunto de matrizes simuladas é o mesmo, portanto a diversidade beta no nível 1 (A) < nível 2 (B) < nível 3 (C) < nível 4 (D), sendo cada nível composto por 30 matrizes.

*Dados simulados de abundância em relação às coordenadas dos gradientes. As matrizes sem correções (dados brutos) tiveram alto ajuste de Procrustes e semelhante aos métodos passo-a-frente (Fig. 7). Os dois métodos passo-a-frente tiveram ajuste*

maior que as transformações de Beals e Swan. Apenas quando o nível de diversidade beta foi alto a transformação de Beals teve desempenho semelhante aos métodos passo-a-frente e sem correções.

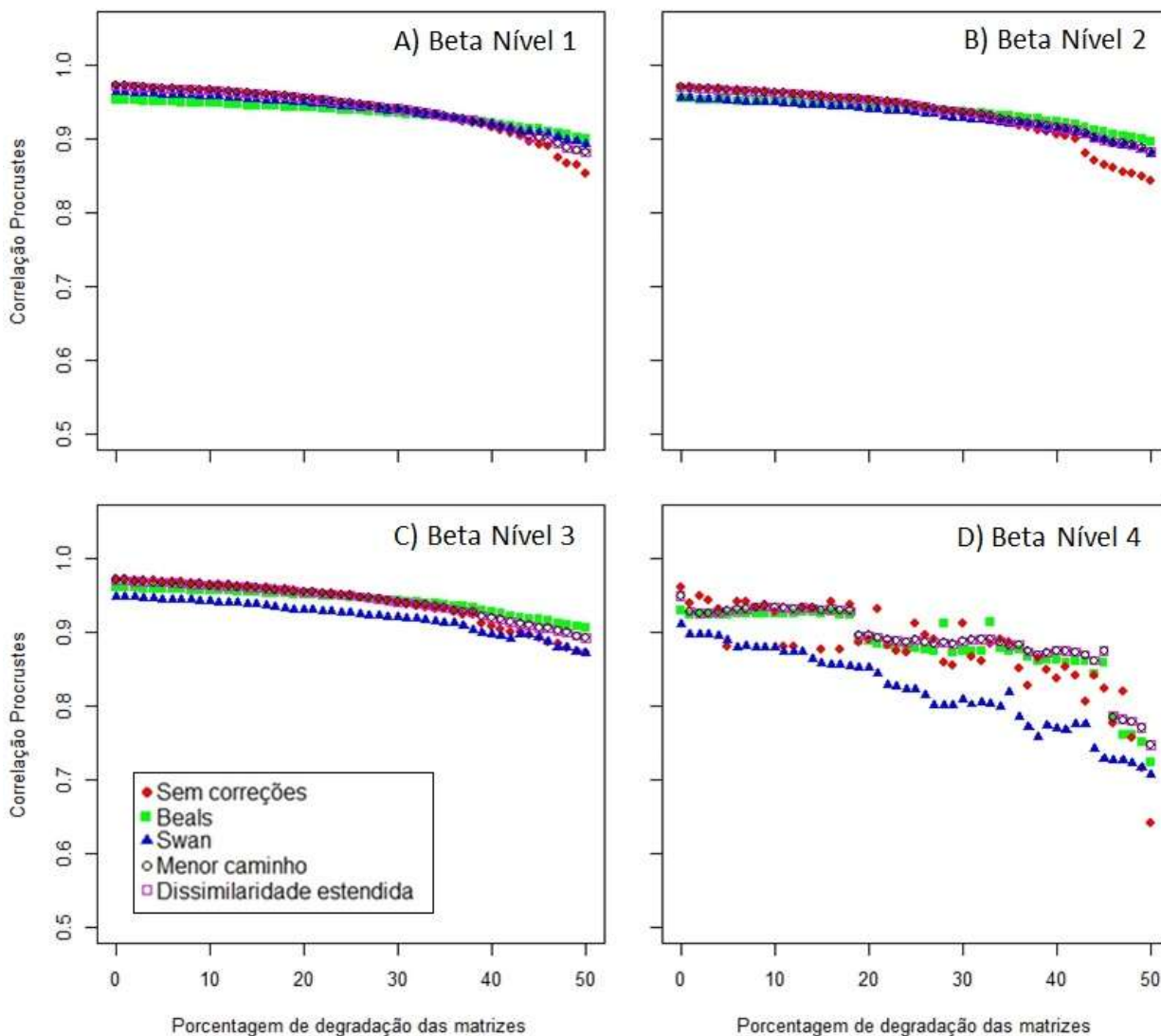


Fig. 8: Média do ajuste do teste Procrustes com 5 formas de medir dissimilaridade (Sorensen, Beals, Swan, Menor Caminho e dissimilaridades estendidas). Utilizei os mesmos conjuntos de dados simulados anteriores em relação às coordenadas das unidades amostrais da simulação. Porém, aqui estão os resultados com as degradações das matrizes de presença ou ausência. No nível 2 (B) estão representadas as 29 matrizes e no nível 4 (D) apenas 3 matrizes, que não tiveram valores ausentes dissimilaridades pelo método passo-a-frente. Os demais diagramas representam 30 matrizes cada.

*Dados simulados de presença-ausência utilizando as coordenadas dos gradientes.* Com a diversidade beta baixa os métodos tiveram ajuste semelhantes, sendo que somente com as matrizes bastante degradadas os métodos de correção apresentaram maior correlação que a análise sem correções (Fig. 8).

*Efeito das propriedades da matriz em relação à inclinação do ajuste ao longo da degradação, com base nas coordenadas da simulação.* As matrizes mais robustas foram as que tinham dados de abundância, maior riqueza total (sem correções), menor proporção de espécies com dois indivíduos, menor proporção de valores zero (exceto com a transformação de Beals), e maior dissimilaridade média (exceto com método de Swan). *Efeito das propriedades da matriz em relação à correlação Procrustes das matrizes sem degradação, com base nas coordenadas da simulação.* Maior ajuste com dados de presença e ausência, menor proporção de espécies com apenas 2 indivíduos, maior dissimilaridade média e menor proporção de dissimilaridades maiores que 0,95 (Tabela 4).

Tabela 4: Resultados da regressão utilizando dados simulados em relação às coordenadas da simulação. Apresento a relação entre a robustez das matrizes à degradação e as suas propriedades. Também apresento a relação do ajuste Procrustes das matrizes sem degradações (correlação da matriz original com a disposição das coordenadas nos gradientes) com as propriedades das matrizes. Estão representados apenas os resultados significativos e se a relação é positiva (+) ou negativa (-).

---

Robustez à degradação

Tratamento dos dados	Tipo de dados	Riqueza total	Proporção de <i>doubletons</i>	Proporção de zeros na matriz	Dissimilaridade média
Brutos	+	+	-	-	+
Beals	+		-		+
Swan	+		-	-	
Menor caminho	+		-	-	+
Diss. Estendidas	+		-	-	+

---

Correlação matriz
sem degradação

Tratamento dos dados	Tipo de dados	Proporção de <i>doubletons</i>	Dissimilaridade média	Proporção de diss. > 0,95
Brutos	-		+	
Beals				-
Swan				
Menor caminho	-	-	+	-
Diss. Estendidas	-	-	+	-

---

#### 4. Discussão

Era esperado que as matrizes degradadas após serem transformadas pelas técnicas de extrapolação das matrizes de espécies ou de dissimilaridades recuperassem melhor o padrão original, tendo em vista que usam informações adicionais (co-ocorrências ou

dissimilaridades com outras unidades amostrais). Os resultados, entretanto, mostraram que os quatro métodos avaliados foram consistentemente piores que a matriz sem transformação. Os diferentes métodos avaliados neste trabalho apresentaram resultados distintos conforme o conjunto de dados utilizado. Com dados empíricos, os métodos passo-a-frente (menor caminho e dissimilaridades estendidas) tiveram maior ajuste do que os métodos de co-ocorrência (Beals e Swan). A transformação de Beals captou bem a ordenação original com dados simulados, porém os métodos passo-a-frente têm maior correlação com as coordenadas das amostras obtidas na simulação. As degradações das amostras têm grande efeito sobre a análise dos gradientes, principalmente com dados empíricos, com os quais encontrei uma grande diminuição no ajuste médio das matrizes sem degradação comparado com as matrizes degradadas. Com dados simulados, o aumento na diversidade beta ampliou a diferença entre os métodos. O padrão dos gradientes ambientais é perdido rapidamente com a degradação de matrizes de presença e ausência. Dados de abundância tendem a ser mais robustos à degradação. Os resultados das análises das propriedades são contrastantes, conforme a origem dos dados e a forma com que eles foram analisados.

De maneira geral, os métodos avaliados distorceram os padrões de semelhança presentes nas matrizes originais. A transformação de Beals é criticada pois pode forçar um padrão na estrutura de dados, mesmo quando ele não existe (De Cáceres & Legendre 2008), o que pode ocorrer com os demais métodos. Desta forma, as informações restantes após a remoção de espécies e indivíduos da matriz, juntamente com as transformações utilizadas não seriam capazes de recuperar o padrão original (ordem das unidades amostrais em um suposto gradiente). Deve-se salientar, entretanto, que nestas avaliações a ordenação referência no Procrustes foi aquela obtida com amostragens de comunidades empíricas ou simuladas. Assume-se, portanto, que estas

ordenações reflitam o padrão real de semelhança das comunidades das quais foram derivadas por amostragem. A avaliação com o uso das coordenadas de simulação como referência no Procrustes, e não das comunidades amostradas nestas coordenadas, apresentou resultados um pouco distintos. Nestas, os métodos apresentaram desempenho semelhante ou levemente superior aos dados das comunidades obtidas nas coordenadas da simulação. O método de dissimilaridades estendidas foi desenvolvido para aumentar a relação das dissimilaridades com a distância ambiental entre as amostras, melhorando a reconstrução do gradiente, principalmente em situações com diversidade beta de moderada a alta (De'ath 1999). Porém, a dissimilaridade de Bray-Curtis têm recebido destaque em estudos ecológicos, visto sua simplicidade, interpretabilidade e bom desempenho em relação a outros índices (Clarke et al. 2006). De fato, os métodos avaliados foram, em geral, piores do que o índice de Bray-Curtis.

Apesar da vantagem de usar dados simulados, pois podemos especificar o padrão antes da análise, é necessário cautela, pois eles nem sempre representam as comunidades reais. Embora dados empíricos obtidos em campo não tenham esta limitação, os padrões verdadeiros na natureza, dos quais eles são derivados, são desconhecidos (Jackson 1993). Assim, meu trabalho pôde avaliar bem a diferença entre dados simulados e empíricos, por que não utiliza variáveis ambientais para avaliar os gradientes empíricos e sim o padrão original da matriz, o que foi feito também com dados simulados. Através de simulações, obtive que quanto menor a diversidade beta os métodos se tornaram mais similares e menos sensíveis à degradação. Assim, quando a diversidade beta é baixa existe pouca diferença entre a medida utilizada ou o esforço amostral, uma vez que nesta situação com remoção de 50 % dos indivíduos das matrizes simuladas o padrão captado não diferiu muito do obtido com as matrizes de abundância sem degradações. Smith (2017) mostrou que o índice de Beals binário possui maior

ajuste do que métodos passo-a-frente e Swan. No meu trabalho este padrão também foi encontrado com dados simulados em relação à ordenação original. Aumentando a diversidade beta, Smith (2017) encontrou que o método passo-a-frente superou a transformação de Beals na reconstrução dos gradientes. Ainda, dados com baixa diversidade beta são robustos à degradação e pouco influenciados pelas dissimilaridades usadas. No meu estudo, com o aumento da diversidade beta os métodos passo-a-frente e a transformação de Beals tiveram alta correlação Procrustes e foram bastante similares. Os índices de Beals e Swan têm forte relação com dados simulados pelo modo com que as matrizes são formadas nas simulações. A estrutura encontrada nas matrizes simuladas é em geral bem clara, o que possivelmente não é encontrado com dados empíricos. Com os dados empíricos, a quantidade de variações aleatórias nas ocorrências de espécies é desconhecida, sugerindo que os métodos passo-a-frente (menor caminho e dissimilaridades estendidas) podem se adaptar melhor às estas incertezas do que os métodos que utilizam probabilidade condicionais com base em co-ocorrências (Beals e Swan). Com relação às coordenadas da simulação, o baixo desempenho utilizando dissimilaridades sem modificações, quando as matrizes foram analisadas com PCoA, também foi encontrado por Smith (2017), porém no meu estudo o ganho com a técnica de nMDS foi tão grande que os métodos se tornam quase indistinguíveis.

Hirst & Jackson (2007) encontraram que técnicas de ordenação com dados de presença e ausência têm melhor desempenho na análise de gradientes do que dados quantitativos, utilizando também o modelo COMPAS para obter as matrizes simuladas. Ou seja, com dados de presença e ausência a estrutura das unidades amostrais foram menos distorcidas do que com dados de abundância. Isto indicaria que a informação de presença ou ausência é capaz de captar a relação entre unidades amostrais em um gradiente, enquanto dados de abundância podem gerar resultados inconsistentes caso

não exista uma relação clara com o ambiente. Avaliando apenas o método, ou seja sem degradações, no meu estudo os dados de presença e ausência tiveram melhor desempenho, medido como maior correlação Procrustes, quando analisado em relação às coordenadas na simulação. A transformação de Beals com abundâncias é pouco aplicada e assim pouco se sabe sobre como o uso das abundâncias para dar pesos ou calcular as propobabilidades condicionais podem afetar os resultados. Austin & Greig-Smith (1968) afirmam que a estrutura da vegetação é um fator que deve ser considerado nos estudos de gradientes, uma vez que a estrutura ecológica que gera o padrão de distribuição das espécies pode não ser óbvia. Assim, as variáveis ambientais não seriam suficientes para predizer a densidade das plantas no local. A abundância de uma espécie em um local pode não estar diretamente relacionada às características ambientais e sim à chegada das espécies, que é, em grande parte, aleatório. Por outro lado, a ocorrência (presença) seria limitada por preferências ou restrições das espécies (Austin & Greig-Smith 1968).

Os dados de abundância são mais robustos à degradação e isso se deve à retirada dos indivíduos ter menor potencial para afetar a estrutura da matriz. Amostras mais abundantes possuem maior chance de perderem um indivíduo, embora ainda mantenham muitos indivíduos naquele local, preservando assim o padrão de ocorrência das espécies. A robustez à degradação também foi explicada pela maior quantidade de espécies na matriz, pois é com base nas ocorrências das espécies que é determinada a ordem das amostras, assim há mais informações para a análise. Com dados empíricos, a maior robustez à degradação foi encontrada nas matrizes com maior proporção de espécies com um ou dois indivíduos, o que pode ser relacionado com outras características destas matrizes. Enquanto que com dados simulados este resultado foi o contrário, o que era esperado, dado que estas espécies (*singletons* e *doubletons*) podem

ser vistas como um ruído neste tipo de análise, pela carência de informações a respeito da relação com as outras espécies. De forma bem consistente, a maior proporção de zeros na matriz está relacionada à menor robustez, uma vez que ela representa a falta de informação para a análise. As matrizes com maior dissimilaridade média foram mais robustas com dados empíricos e com dados simulados em relação às coordenadas, o que era inesperado, posto que estas dissimilaridades devem aumentar ainda mais com a retirada de indivíduos, distorcendo ainda mais a ordenação. Com os dados simulados em relação à ordenação original, as matrizes com maior dissimilaridade média foram mais robustas, no entanto aquelas com maior proporção de dissimilaridade maiores que 0,95 foram mais robustas, o que novamente pode estar relacionado com outros fatores, por exemplo, com o tamanho da matriz.

Avaliando apenas os métodos, ou seja sem degradação da matriz, sobre os dados empíricos as matrizes com maior correlação Procrustes foram as de abundância, com maior quantidade de unidades amostrais e maior proporção de espécies com uma ocorrência. Além disso, as matrizes empíricas com menor proporção de zeros, menor dissimilaridade média e menor proporção de dissimilaridades maiores que 0,95 tiveram maior ajuste Procrustes, pois estas matrizes fornecem informações mais completas. Com dados simulados em relação às coordenadas, temos que dados de presença e ausência e matrizes com menor proporção de espécies com dois indivíduos têm maior ajuste Procrustes, conforme esperado. Além disso, os dados simulados com maior dissimilaridade média e com a menor proporção de dissimilaridades maiores que 0,95 tiveram maior correlação Procrustes. Isto também é inesperado, pois matrizes com menor proporção de dissimilaridade maiores que 0,95 devem ter menor dissimilaridade média.

A análise multivariada de comunidades se desenvolveu muito desde trabalhos pioneiros e ainda avança rapidamente, embora ainda existam discordâncias entre as estratégias mais efetivas para determinadas propostas (Minchin e Oksanen 2015). No meu trabalho criei valores ausentes propositalmente para testar se as estratégias utilizadas para correção de dados recuperavam o padrão original. Estudos sobre valores ausentes recebem pouca atenção, sendo uma grande oportunidade para pesquisas em Ecologia para desenvolvimento analítico das técnicas para suplementar a falta de dados (Dray e Josse 2015). Meu trabalho mostra que nos casos em que o padrão de similaridades entre unidades amostrais é destruído (através da perda de esforço amostral) as extrapolações utilizadas não reconstróem a relação que existia, como também deforma ainda mais o padrão original. Assim, as transformações e extrapolações devem ser usadas com cuidado. É bom que haja como avaliar os processos ecológicos que geram o padrão de diversidade observado e garantir que a relação entre as amostras não é espúria.

## 5. Referências

- Austin, M. P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio*, 69: 35-45.
- Austin, M. P. 2013. Inconsistencies between theory and methodology: a recurrent problem in ordination studies. *Journal of Vegetation Science*, 24(2): 251-268.
- Austin, M. P., & Greig-Smith, P. 1968. The application of quantitative methods to vegetation survey: II. Some methodological problems of data from rain forest. *The Journal of Ecology*, 56(3): 827-844.
- Atmar, W. & B.D. Patterson. 1995. The nestedness temperature calculator: a visual basic program, including 294 presence-absence matrices. AICS Research, Inc., University Park, NM, and The Field Museum, Chicago. Disponível em: <<https://sites.google.com/a/fieldmuseum.org/bruce-pattersons-lab/Home/>>

research-areas/nested-subsets>, Baixado 9 de outubro de 2017.

- Clarke, K. R., & Ainsworth, M. (1993). A method of linking multivariate community. *Marine ecology progress series*, 92: 205-219.
- Clarke, K. R., Somerfield, P. J., & Chapman, M. G. (2006). On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *Journal of Experimental Marine Biology and Ecology*, 330(1): 55-80.
- De'ath, G. (1999). Extended dissimilarity: a method of robust estimation of ecological distances from high beta diversity data. *Plant Ecology*, 144(2): 191-199.
- De Cáceres, M., & Legendre, P. (2008). Beals smoothing revisited. *Oecologia*, 156(3): 657-669.
- Dray, S., & Josse, J. (2015). Principal component analysis with missing values: a comparative survey of methods. *Plant Ecology*, 216(5): 657-667.
- Faith, D. P., Minchin, P. R., & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69(1-3): 57-68.
- Hirst, C. N., & Jackson, D. A. (2007). Reconstructing community relationships: the impact of sampling error, ordination approach, and gradient length. *Diversity and Distributions*, 13(4): 361-371.
- Jackson, D. A. (1993). Multivariate analysis of benthic invertebrate communities: the implication of choosing particular data standardizations, measures of association, and ordination methods. *Hydrobiologia*, 268(1): 9.
- Jackson, D. A., Somers, K. M., & Harvey, H. H. (1989). Similarity coefficients: measures of co-occurrence and association or simply measures of occurrence?. *The American Naturalist*, 133(3): 436-453.
- McCune, B. (1994). Improving community analysis with the Beals smoothing function. *Ecoscience*, 1(1): 82-86.
- Melo, A. S. 2016. Community ecology analyses. – R package vers. 1.5.9 < [www.r-project.org](http://www.r-project.org) >.
- Minchin, P. R. 1987a. An evaluation of relative robustness of techniques for ecological ordinations. *Vegetatio*, 69: 89-107.
- Minchin, P. R. 1987b. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio*, 71: 145-156.
- Minchin, P. R., & Oksanen, J. (2015). Statistical analysis of ecological communities: progress, status, and future directions. *Plant Ecology*, 216(5): 641-644.
- Münzbergová, Z. & T. Herben. 2004. Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos* 105: 408-414.

- Oksanen, J. & P. R. Minchin. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, 157: 119-129.
- Oksanen, J., Blanchet F., Friendly M., Kindt R., Legendre P., McGlinn D., ... & Wagner H. (2017). vegan: Community Ecology Package. R package version 2.4-2. <https://CRAN.R-project.org/package=vegan>
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Smith, R. J. 2017. Solutions for loss of information in high-beta-diversity community data. *Methods in Ecology and Evolution*, 8(1): 68-74.

## Material Suplementar (I)

Fonte e descrição das matrizes de dados empíricos de abundância obtidas no repositório Dryad.

### I.1) Braton;

Grupo: Invertebrados

Classificação: Família

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.r762t>

Fonte: Branton MA, Richardson JS (2014) A test of the umbrella species approach in restored floodplain ponds. *Journal of Applied Ecology* 51(3): 776–785. <https://doi.org/10.1111/1365-2664.12248>

Amostragem: Os invertebrados das lagoas foram amostrados utilizando redes. Padronizado por tempo

### I.2) Rudolf; I.3) Macro\_invertebrateSubsampleWide; I.4) Zooplankton;

### I.5) Final.MacroInverts;

Grupos: macroinvertebrados ou zooplancton

Classificação: outros

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.5bm68>

Fonte: Rudolf VHW, Van Allen BG (2017) Legacy effects of developmental stages determine the functional role of predators. *Nature Ecology & Evolution* 1: 0038. <https://doi.org/10.1038/s41559-016-0038>

Amostragem: Todos os macro-invertebrados foram identificados, contados e retornaram para os tanques durante o experimento exceto pequenos invertebrados que foram fixados para identificação. Amostras de zooplâncton nas bordas e no centro de cada tanque (volume total: 2,5 l) obtidas com um tubo amostrador. As subamostragens foram realizadas retirando os macroinvertebrados do tanque com uma rede durante o experimento. Ao final todos os indivíduos foram retirados do tanque e identificados (Final.MacroInverts)

### I.6) Satdichanh;

Grupo: árvores

Classificação: espécies

Baixado em: <http://datadryad.org/handle/10255/dryad.88632>

Fonte: Satdichanh M, Millet J, Heinemann A, Nanthavong K, Harrison RD (2015) Using plant functional traits and phylogenies to understand patterns of plant

community assembly in a seasonal tropical forest in Lao PDR. PLOS ONE 10(6): e0130151. <https://doi.org/10.1371/journal.pone.0130151> 11

Amostragem: parcelas de 50x50 m com todas árvores com mais 10 cm de DAP (diâmetro acima do peito).

I.7) Theuerkauf1; I.8) Theuerkauf2;

Grupo: invertebrados marinhos

Classificação: gênero/espécies

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.5qp96/3>

Fonte: Theuerkauf KW, Eggleston DB, Theuerkauf SJ (2017) An exotic species alters patterns of marine community development. *Ecological Monographs.* , 88(1), 92-108. <https://doi.org/10.1002/ecm.1277>

Amostragem: A estrutura das comunidades a curto prazo através da colonização das larvas dos invertebrados foram obtidas em 3 pratos submersos de 232 cm<sup>2</sup>, amostrados mensalmente. A estrutura a longo prazo é obtida ao final do experimento nas observações de todas as réplicas ao longo do tempo.

I.9) utsumi;

Grupo: artrópodes

Classificação: gênero/espécies

<http://datadryad.org/resource/doi:10.5061/dryad.s2nr5>

Fonte: Utsumi S (2015) Feeding evolution of a herbivore influences an arthropod community through plants: implications for plant-mediated eco-evolutionary feedback loop. *Journal of Ecology* 103(4): 829-839. <https://doi.org/10.1111/1365-2745.12419>

Amostragem: Os invertebrados foram coletados em árvores. Todas as plantas do experimento foram dispostas em quatro blocos e ajustadas aleatoriamente em uma grade de quatro por cinco linhas com espaçamento de 1 m em cada bloco. Cada bloco continha quatro plantas de cada árvore de origem. A cada planta foi atribuído um dos dois tratamentos e as demais foram definidas como controle.

I.10) zhu;

Grupo: plantas herbáceas

Classificação: espécies

Baixado em: <http://datadryad.org/handle/10255/dryad.84628>

Fonte: Zhu H, Fu B, Wang S, Zhu L, Zhang L, Jiao L, Wang C (2015) Reducing soil erosion by improving community functional diversity in semi-arid grasslands. *Journal of Applied Ecology* 52(4): 1063-1072. <http://dx.doi.org/10.1111/1365-2664.12442>

Amostragem: Dezesesseis comunidades herbáceas foram selecionadas para o levantamento de vegetação. Adjacente a cada parcela, foram estabelecidas cinco parcelas de 1 × 1 m. Os valores médios obtidos nas cinco parcelas foram utilizados para representar a composição da vegetação da parcela. Todas as espécies de plantas presentes em cada parcela foram identificadas.

I.11) rich;

Grupo: mamíferos

Classificação: espécies

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.q54rp>

Fonte: Rich LN, Miller DA, Robinson HS, McNutt JW, Kelly MJ (2016) Using camera trapping and hierarchical occupancy modelling to evaluate the spatial ecology of an African mammal community. *Journal of Applied Ecology* 53(4): 1225-1235. <https://doi.org/10.1111/1365-2664.12650>

Amostragem: Detecção feita por 220 armadilhas fotográficas.

I.12) helsen;

Grupo: plantas (traqueófitas)

Classificação: espécies

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.7s5s4>

Fonte: Helsen K, Hermy M, Honnay O (2016) A test of priority effect persistence in semi-natural grasslands through the removal of plant functional groups during community assembly. *BMC Ecology* 16: 22. <https://doi.org/10.1186/s12898-016-0077-9>

Amostragem: 48 parcelas de 5 × 5 m.

I.13) valtonen;

Grupo: lepidópteros

Classificação: espécies

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.9m6vp>

Fonte: Valtonen A, Hirka A, Szócs L, Ayres MP, Roininen H, Csóka G (2017) Long-term species loss and homogenization of moth communities in Central Europe. *Journal of Animal Ecology* 86(4): 730-738. <https://doi.org/10.1111/1365-2656.12687>

Amostragem: diversas.

I.14) patrickAll\_fiftykmcircles; I.15) patrickAll\_fivekmbuffer;  
 I.16) patrickAll\_tenkmbuffer; I.17) patrickCH\_fiftykmcircles;  
 I.18) patrickCH\_fivekmbuffer; I.19) patrickCH\_tenkmbuffer;  
 I.20) patrickGB\_fiftykmcircles; I.21) patrickGB\_fivekmbuffer;  
 I.22) patrickGB\_tenkmbuffer; I.23) patrickMJ\_fiftykmcircles;  
 I.24) patrickMJ\_fivekmbuffer; I.25) patrickMJ\_tenkmbuffer;  
 I.26) patrickSN\_fiftykmcircles; I.27) patrickSN\_fivekmbuffer;  
 I.28) patrickSN\_tenkmbuffer; I.29) patrickAll\_fiftykmgrid;  
 I.30) patrickAll\_hundredkmcircles; I.31) patrickAll\_tenkmgrid;  
 I.32) patrickCH\_fiftykmgrid; I.33) patrickCH\_hundredkmcircles;  
 I.34) patrickCH\_tenkmgrid; I.35) patrickGB\_fiftykmgrid;  
 I.36) patrickGB\_hundredkmcircles; I.37) patrickGB\_tenkmgrid;  
 I.38) patrickMJ\_fiftykmgrid; I.39) patrickMJ\_hundredkmcircles;  
 I.40) patrickMJ\_tenkmgrid; I.41) patrickSN\_fiftykmgrid;  
 I.42) patrickSN\_hundredkmcircles; I.43) patrickSN\_tenkmgrid;

Grupo: morcegos

Classificação: espécies

Baixado em: <http://datadryad.org/resource/doi:10.5061/dryad.dq383>

Fonte: Patrick LE, Stevens RD (2016) Phylogenetic community structure of North American desert bats: influence of environment at multiple spatial and taxonomic scales. *Journal of Animal Ecology* 85(4): 1118–1130. <https://doi.org/10.1111/1365-2656.12529>

Amostragem: Registros de campo próprios e outros registros tais como museus\*

\*As 30 matrizes cujos nomes iniciam com Patrick compõem 6 delineamentos diferentes (*buffer* 5 quilômetros: “fivekmbuffer”; *buffer* 10 quilômetros: “tenkmbuffer”; grade de 5 quilômetros: “fivekmgrid”; grade de 10 quilômetros: “tenkmgrids”; círculos de 50 quilômetros: “fiftykmcircles” e 100 quilômetros: “hundredkmcircles”. Em 4 desertos diferentes (GB, MJ, SN e CH) e o conjunto de todos os desertos (ALL).

## Material Suplementar (II)

Correlação de Procrustes médio com a ordenação nMDS para os dados de abundância e com a técnica PCoA com os dados simulados.

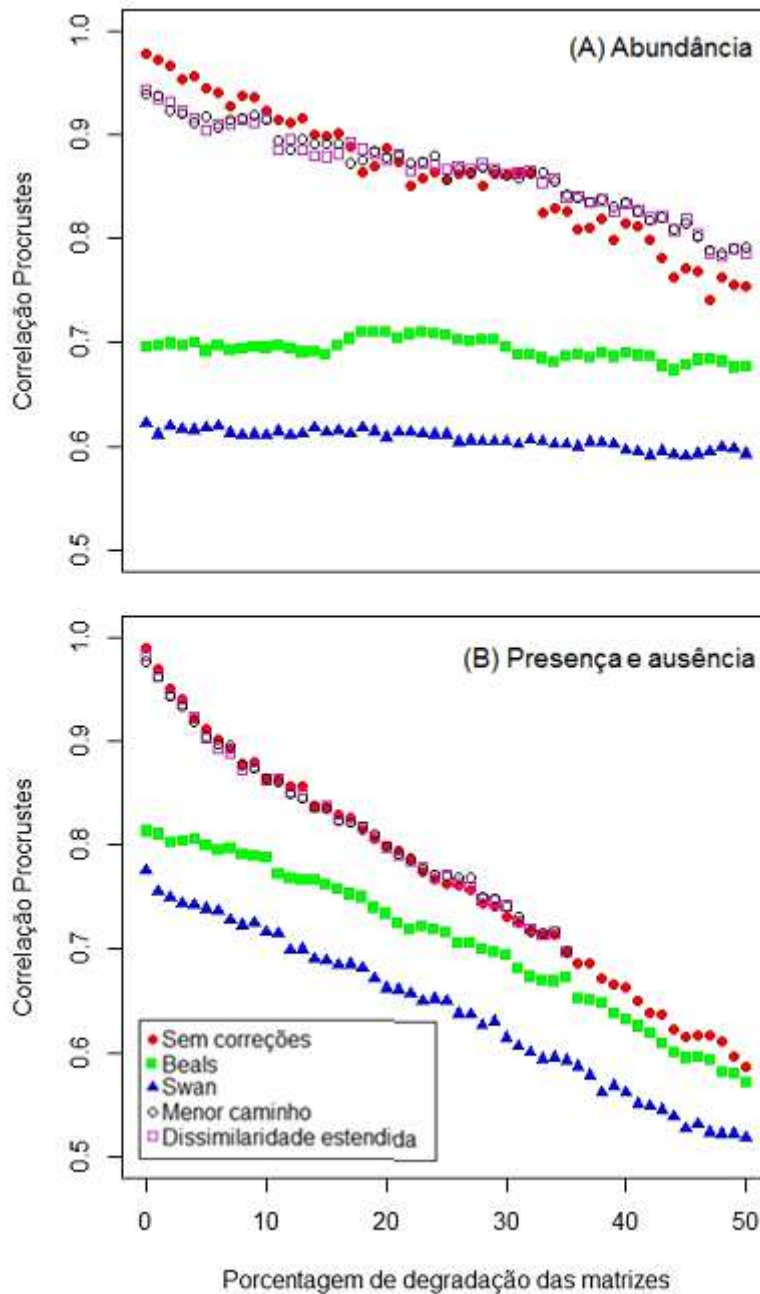


Fig. 1 (MS): Média da correlação de Procrustes com dados empíricos usando a técnica de ordenação nMDS. Os dados de abundância (A) representam as médias dos ajustes de 40 matrizes empíricas e os dados de presença e ausência (B) são compostos de 140

matrizes originalmente de presença e ausência e 17 matrizes transformadas para presença e ausência das 40 matrizes de abundância.

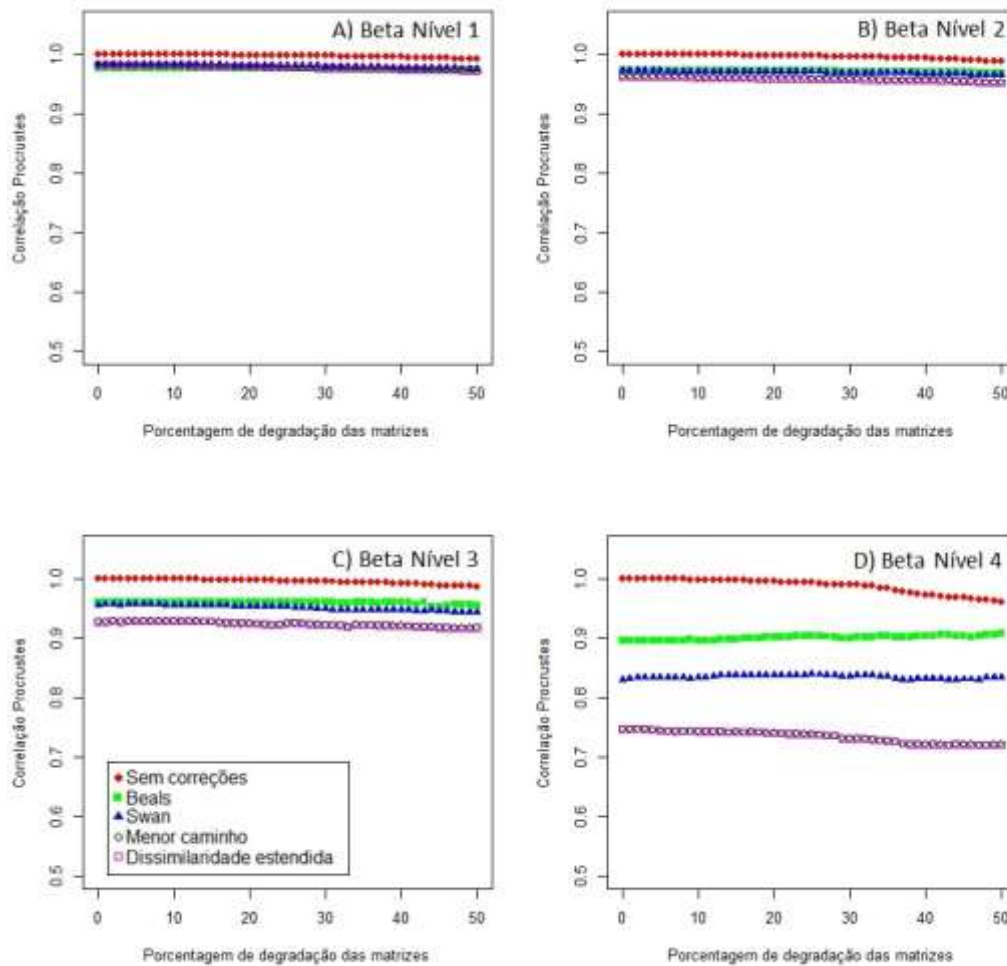


Fig. 2 (MS): Média da correlação de Procrustes com dados simulados de abundância usando a ordenação PCoA. Com 30 matrizes em cada nível de diversidade beta (determinado pelo parâmetro beta do modelo), sendo: nível 1 (A) com as matrizes cuja diversidade beta é menor até o nível 4 (D) com a maior diversidade beta.

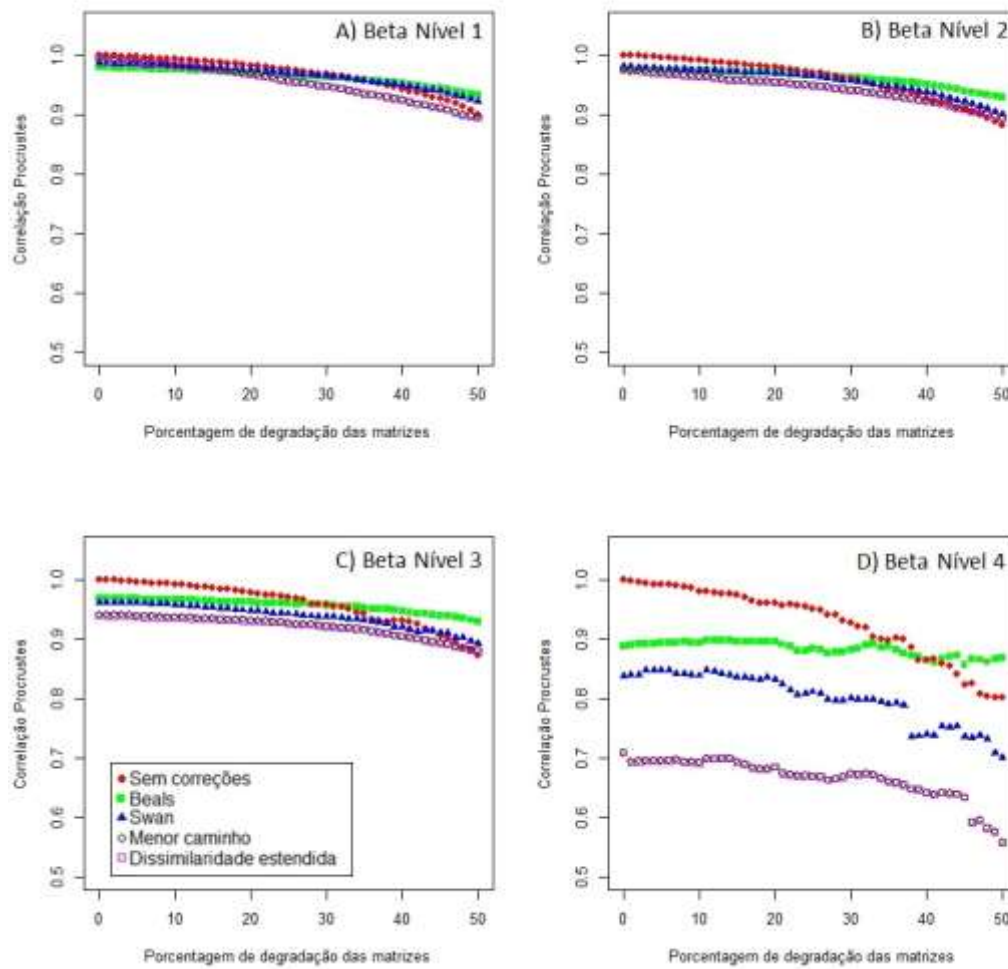


Fig. 3 (MS): Média da correlação de Procrustes com dados simulados de presença e ausência usando a ordenação PCoA. Como as matrizes utilizadas são as mesmas da figura anterior, os níveis de diversidade beta continuam os mesmos. Porém no nível 2 (B) uma matriz foi retirada, enquanto no nível 4 (D) restaram apenas 3 matrizes sem valores ausentes na dissimilaridades. Desta forma estes níveis representam um subconjunto das suas respectivas matrizes de abundância.

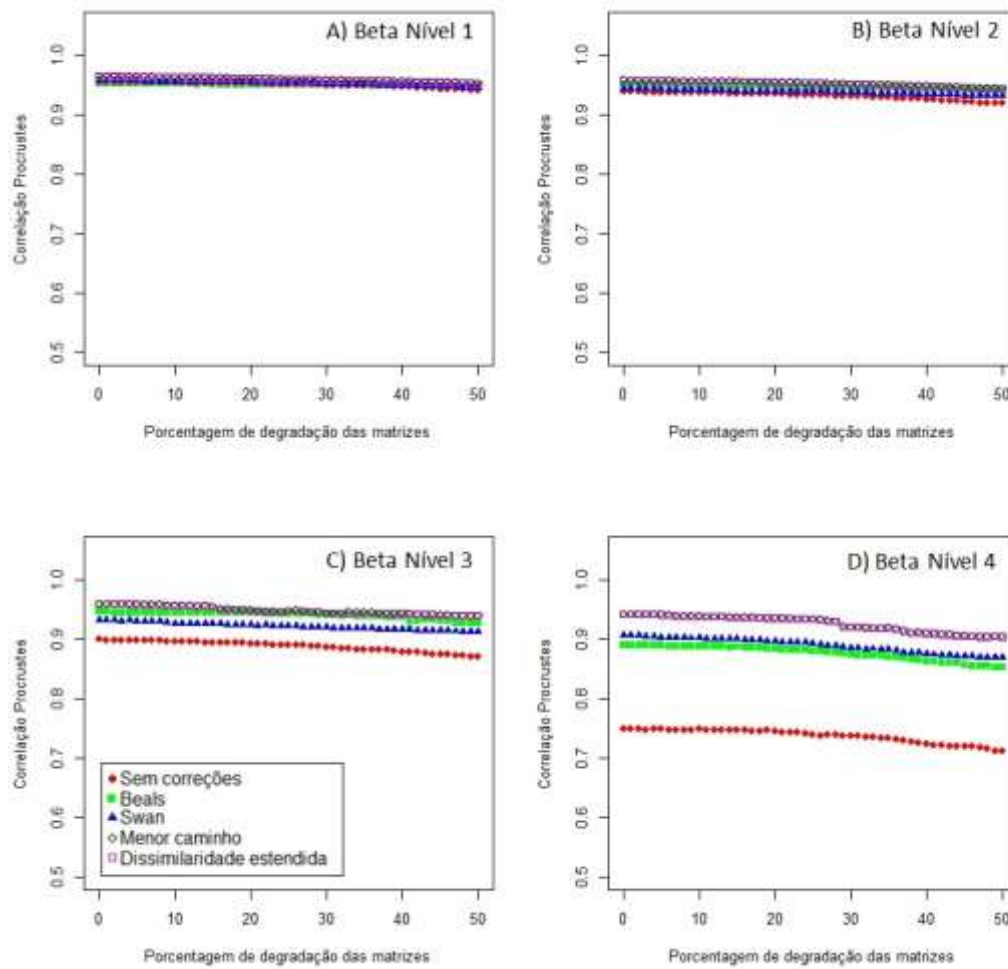


Fig. 4 (MS): Média da correlação de Procrustes com dados simulados de abundância usando a ordenação PCoA em relação às coordenadas das amostras. Com 30 matrizes em cada nível de diversidade beta (determinado pelo parâmetro beta do modelo), sendo: nível 1 (A) com as matrizes cuja diversidade beta é menor até o nível 4 (D) com a maior diversidade beta.

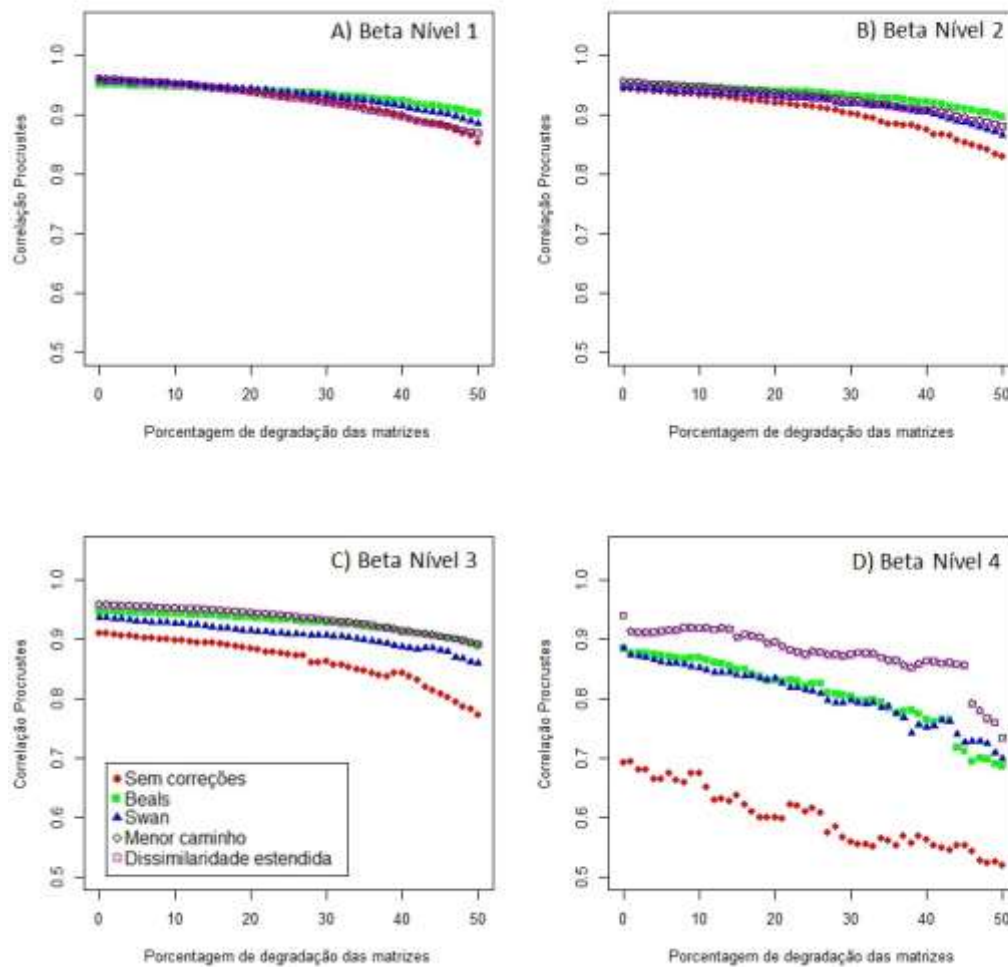


Fig. 5 (MS): Média da correlação de Procrustes com dados simulados de presença e ausência usando a ordenação PCoA em relação às coordenadas das amostras. Utilizei os mesmos conjuntos de dados simulados anteriores, porém, aqui estão os resultados com as degradações das matrizes de presença ou ausência. No nível 2 (B) estão representadas as 29 matrizes e no nível 4 (D) 3 matrizes, que não tiveram valores ausentes dissimilaridades pelo método passo-a-frente. Os demais diagramas representam 30 matrizes cada.

**Observação:** Para acesso às matrizes de dados e códigos implementados no R requisi-teo ao autor por e-mail no endereço [j.cury.h@gmail.com](mailto:j.cury.h@gmail.com)