UFRRJ

INSTITUTO DE AGRONOMIA

CURSO DE PÓS-GRADUAÇÃO EM AGRONOMIA CIÊNCIA DO SOLO

TESE

Uso da Abordagem Estatística Procrusteana em Ecologia de Solo: Caso de Estudo Envolvendo Sistema de Integração Lavoura-Pecuária-Floresta no Cerrado

Francy Junio Gonçalves Lisboa

2015

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
CURSO DE PÓS-GRADUAÇÃO EM AGRONOMIA
CIÊNCIA DO SOLO

USO DA ABORDAGEM ESTATÍSTICA PROCRUSTEANA EM ECOLOGIA DE SOLO: CASO DE ESTUDO ENVOLVENDO SISTEMA DE INTEGRAÇÃO LAVOURA-PECUÁRIA-FLORESTA NO CERRADO

FRANCY JUNIO GONÇALVES LISBOA

Sob orientação do Professor
Ricardo Luis Louro Berbara

Co-orientação
Guilherme Montandon Chaer

Tese submetida como requisito parcial para obtenção do grau de **Doutor,** no Curso de Pós-Graduação em Agronomia, Área de Concentração em Ciência do Solo

Seropédica, RJ
Fevereiro de 2015

Ficha catalográfica

634.99
L769u
T

Lisboa, Francy Junio Gonçalves, 1985-
Uso da abordagem estatística procrusteana em ecologia de solo: caso de estudo envolvendo sistema de integração lavoura-pecuária-floresta no cerrado / Francy Junio Gonçalves Lisboa. – 2015.
90 f.: il.

Orientador: Ricardo Luis Louro Berbara.
Tese (doutorado) – Universidade Federal Rural do Rio de Janeiro, Curso de Pós-Graduação em Agronomia – Ciência do Solo, 2015.
Inclui bibliografia.

1. Agrossilvicultura - Teses. 2. Ecologia do solo – Métodos estatísticos - Teses. 3. Solo - Uso - Teses. 4. Micro-organismos do solo – Teses. 5. Plantas e solo – Teses. 6. Análise multivariada – Teses. I. Berbara, Ricardo Luis Louro, 1957- II. Universidade Federal Rural do Rio de Janeiro. Curso de Pós-Graduação em Agronomia – Ciência do Solo. III. Título.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO DE AGRONOMIA
CURSO DE PÓS-GRADUAÇÃO EM AGRONOMIA-CIÊNCIA DO SOLO

FRANCY JUNIO GONÇALVES LISBOA

Tese submetida como requisito parcial para obtenção do grau de **Doutor**, no Curso de Pós-Graduação em Agronomia, área de Concentração em Ciência do Solo.
.

TESE APROVADA EM 25/02/2015

_____
Ricardo Luis Louro Berbara. Ph.D. UFRRJ
(Orientador)

_____
Marcos Gervasio Pereira. Dr. UFRRJ

_____
Gustavo Ribeiro Xavier. Dr. Embrapa Agrobiologia

_____
Janaína Ribeiro da Costa Rouws. Dra. Embrapa Agrobiologia

_____
Beata Emoki Madari. Ph.D. Embrapa Arroz e Feijão

Ao meu filho Miguel Lima Lisboa, e à minha companheira Mariana Gomes Lima

Dedico

AGRADECIMENTOS

Gostaria de agradecer a todos aqueles que contribuíram, seja de forma direta ou indireta, para que a tese fosse configurada. Agradeço à minha família: Mariana Lima e Miguel Lisboa simplesmente por existirem, servindo de refúgio contra qualquer contrapeso externo. Ao meu cavaquinho por amansar momentos de cólera. Apesar de ainda tentar me converte para a erudição da música clássica, agradeço ao meu parceiro e amigo, Ederson Conceição Jesus, pelos debates de ciência, sobre a ciência, e aqueles nem tão científicos assim. Por falar em Ederson, não poderia deixa de agradecer ao laboratório de leguminosas, desde os bolsistas (Felipe Martini, Khadidja Dantas, Alessandro (flecha), Danilo Ataíde, Joel Quintino) até os "caciques", como Alexander Resende, Sérgio Miana (até hoje dizendo pra Deus emundo que neguei emprego na Vale), Eduardo Campelo, Guilherme Chaer (co-orientador), Luis Fernando, e Juliana Muller . Ainda, agradeço a Andreia Cunha pela amizade e companheirismo. Contudo, minha singela homenagem ao saudoso **Telmo Félix** e ao Fernando Cunha (o pastor), dois dos principais motivos para tornar a vida no laboratório de leguminosa mais animada. Agradeço ao doutor Marcelo Ferreira Fernandes por abrir as portas do seu laboratório na Embrapa Aracajú para que as análises de ácidos graxos, cerne da presente tese, pudessem ser conseguidas. Agradeço à Dra. Beata Madari e a **Capes pela concessão da bolsa**. Agradeço à Dra. Beata e ao **Curso de Pós – graduação em Agronomia - Ciência do Solo** por todo aporte para a configuração da presente tese.

Saindo da seara embrapiana, meus agradecimentos são dirigidos ao grupo do laboratório de biologia do Solo da UFRRJ. Agradeço a Andres Calderín pelas conversas sobre ciência, artigos, etc. Não poderia deixar, obviamente de agradecer ao orientador e amigo, Ricardo Berbara, por deixar não deixar que as linhas da natural hierarquia existente entre orientado e orientador dificultassem o meu processo de liberdade para propor e compor ciência; agradeço também por ser sempre solista ao buscar soluções para ajudar não só a mim, mas todos no laboratório de Biologia do Solo.

Por fim, gostaria de agradecer aos parceiros estrangeiros que obtive durante o período de doutorado sanduíche no The James Hutton Institute (Aberdeen – Escócia – Reino Unido). Superando as expectativas, agradeço por ter sido tão bem recebido por toda a equipe do Hutton. Em especial, agradeço aos meus orientadores Ruth Joy Mitchell e Steve James Chapman pela paciência ímpar com meu inglês ao atinarem que a parceria científica poderia ser producente para todos independente da proficiência. Agradeço aos dois também pela incrível hospitalidade ao abrirem as portas de suas casas para mim; e agradeço ainda mais por perceber que a parceria foi solidificada a ponto de durar além do período que estive no Hutton. No âmbito da vida cotidiana na Escócia, eu agradeço, e muito, ao meu ex *flatmate*, Sam Gandy, pela paciência, além de tornar a vida longe de esposa e filho um pouco menos triste.

BIOGRAFIA

Francy Lisboa nasceu em 03/07/1985 em Angra dos Reis. Ingressou no curso de Agronomia em 2005 e até 2006 não havia atinado para o mundo científico. Somente em 2007 tornou-se estudante de iniciação científica sob a orientação do professor Maurício Balesteiro, trabalhando com tomate cereja. Apesar de agradecer pela primeira oportunidade, essa não era o seudestino. A oportunidade para trabalhar com ecologia de solo surgiu no final de 2007 depois de ser aprovado na prova da Embrapa, e ser aceito pelo então pesquisador da Embrapa Agrobiologia, Francisco Adriano de Souza. Na Embrapa Agrobiologia teve o primeiro contato com a microbiologia do solo, especificamente, Fungos Micorrízicos Arbusculares e, por tabela com seu atual orientador, Ricardo Berbara. Com a saída de Francisco Adriano da unidade da Embrapa no final de 2009, tornou-se estudante de iniciação no laboratório de leguminosas florestas sob a orientação do pesquisador Sérgio Miana. Em março de 2010 ingressou no Curso de Pós-Graduação em Agronomia-Ciência do Solo (CPGA-CS) sob a orientação do professor Ricardo Berbara. Contudo, seu projeto foi conduzido nas minas de Carajás e financiado pela empresa Vale S.A por intermédio de Sérgio Miana. Os dois anos de projeto culminaram na dissertação: **"Vínculos entre Variáveis de Solo, Plantas e Ambiente em Áreas Revegetadas após Mineração na Amazônia"**. Em março de 2012 ingressou no doutorado no mesmo CPGA-CS, mas agora trabalhando em projeto multinstitucional denominado CARBIOM: "agricultura sustentável no bioma Cerrado", e coordenado pela pesquisadora Beata Emoki Madari (Embrapa Arroz e Feijão). Dentro das múltiplas atribuições dentro do CARBIOMA, Francy Lisboa ficou responsável por acessar as variações da estrutura microbiana em sistema de integração lavoura-pecuária-floresta. Os resultados da mais atual empreitada são apresentados na presente tese intitulada: "**Uso da Abordagem Estatística Procrusteana em Ecologia de Solo: caso de estudo envolvendo sistema de integração lavoura-pecuária-floresta no Cerrado**".

RESUMO GERAL

LISBOA, Francy Junio Gonçalves. **Uso da abordagem estatística procrusteana em ecologia de solo: caso de estudo envolvendo sistema de integração lavoura-pecuária-floresta no Cerrado**. 2015. 90f. Tese (Doutorado em Agronomia - Ciência do Solo). Instituto de Agronomia, Departamento de Solos, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2015.

A presente tese fez parte do esforço multinstitucional buscando sustentar a substituição de pastagens degradas por sistemas que integrem diferentes tipos de uso da terra, mais especificamente aqueles integrando lavoura, pastagem, e floresta plantada, coletivamente: sistemas iLPF. Aqui, o foco foi a exploração das potencialidades da abordagem estatística denominada análise Procrutes, ou simplesmente Procrustes, na seara de ecologia de planta e solo. Basicamente, a tese foi composta por três capítulos onde é descrito com detalhes os principais nuances dessa abordagem multivariada ainda pouco utilizada por ecologistas de planta e solo. O primeiro capítulo descreve roteiros esquemáticos mostrando como o vetor de resíduos derivado da correlação e duas tabelas de dados pela análise Procrustes (chamado PAM: *Procrustes association metric*) pode ser utilizado como representante univariado da correlação em outras abordagens estatísticas (ordenação ecológica, regressão, e ANOVA seguida de teste de médias). O segundo capítulo da tese, utilizando sugestões do primeiro capítulo, tratou de um estudo de caso. Neste caso, fazenda experimental situada no município de Cachoeira dourada – GO, e contendo quatro diferentes tipos de uso da terra, dentre os quais um sistema iLPF, foi escolhida para a condução do estudo de caso. O objetivo geral foi acessar como correlações, no formato de PAM, entre tabelas de dados representadas por variáveis individuais de estrutura microbiana (dada por análise de lipídios oriundos do solo; PLFA: *Phospholipids Fatty Acid*) e propriedades individuais de química e física de solo, eram moduladas pelo tipo de uso da terra: pastagem degradada, pastagem melhorada, fragmento de mata nativa, e sistema iLPF. A hipótese para o estudo de caso foi a de que a relação fungo: bactéria, comumente associada a ambientes mais conservativos, era promovida pelo sistema iLPF uma vez que tais sistemas são caracterizados pelo aumento da heterogeneidade vegetal oriunda da sistematizada introdução de especies arbóreas em meio a pastagem. O terceiro e último capítulo da tese foi estritamente dedicado a responder questionamentos técnicos referentes à abordagem procrusteana e surgidos depois das publicações dos dois primeiros capítulos da tese. Neste caso, dois dos questionamentos mais comuns foram abordados. Foram eles: i) quais são os efeitos da correlação entre colunas/variáveis dentro de uma tabela de dados sobre os resultados da análise Procrustes? ii) Pode o vetor de resíduos procrusteanos, a PAM, traduzir diferenças entre tratamentos em termos da força de correlação multivariada entre duas tabelas de dados? Para o estudo de caso os resultados da corrente tese suportaram os sistemas iLPF como potencial alternativa para substituição de pastagens degradadas ao levantar indícios de que a heterogeneidade vegetal introduzida nos sistemas iLPF pode favorecer o deslocamento da estrutura microbiana em direção ao domínio de fungos.

Palavras-chave:ILPF. Estrutura Microbiana do Solo. PLFA. Análise Multivariada. Relação Fungo:Bactéria

GENERAL ABSTRACT

LISBOA, Francy Junio Gonçalves. **Uses of the Procrustean statistical approach in soil ecology: a case of study involving an integrated agroecosystem in Brazilian savannah.** 2015.90p. Thesis (Doctor Science in Agronomy, Soil Science). Instituto de Agronomia, Departamento de Solos, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2015.

This thesis is part of a multiple scientific effort seeking to support the replacement of degraded brazilian pastures by systems which integrate different land use types such as crop, pasture, and forest plantation (collectively known as iCLF systems). Here, the focus was also to discuss the potentialities of an unusual statistical multivariate approach called "Procrustes Analysis" in the plant and soil ecology framework. The current thesis has three chapters through which details of the Procrustes analysis are presented on both technically e intuitively manner. The first chapter describes roadmaps showing how the procrustean residual vector (so-called PAM: *Procrustean association metric*), representing the multivariate correlation between two or more data tables, can be used as an univariate variable in more user-traditional statistical approaches such as ecological ordination, regression analysis and ANOVA followed by mean comparisons. The second chapter discussed a case study and had as the general objective to use PAMs, depicting the relationships between distance matrices from individual soil microbial structure (PLFA: *Phospholipids Fatty Acid*) and distance matrices form soil properties variables (chemical and physic), as response variables in an ANOVA framework with land use type as categorical predictor (degraded pasture, improved pasture, native fragment and iCLF system). The hypothesis in this case was that the fungi:bacteria ratio given by PLFA analysis, a good index of changes in microbial structure as response to land use alteration and associated to more conservative soils in terms of carbon mineralization, is favored by the man – introduced vegetal heterogeneity which characterizes the integration crop – livestock – forest. The last chapter was entirely dedicated to answer some technical questions which arose after the publication of the first chapters. Basically the two most common questions were: **i**) Does the increasing number of columns/variables within a data table affect Procrustes outcomes? **ii**) Can the procrustean residual vector, the PAM, translate differences between treatments in terms of multivariate correlation as it is used in mean comparisons? Specifically for these questions, Procrustes was useful in supporting iCLF systems as potential alternative to degraded pasture by raising insights that the man – introduced vegetal heterogeneity in such integrated agroecosystem, favor shifts in microbial structure toward fungal dominance.

Keywords: iCLF. Soil microbial structure. PLFA. Multivariate analysis. Fungi: Bacteria ratio.

SUMÁRIO

1. INTRODUÇÃO GERAL

No dia 17 de agosto de 2010 a portaria 3.896 emitida pela autoridade monetária brasileira, o Banco Central, estabeleceu nova linha de crédito direcionada exclusivamente ao financiamento de projetos voltados à sustentabilidade agropecuária. Essa linha de crédito, por convenção, foi chamada Programa de Agricultura de Baixa Emissão de Gases do efeito estufa, mais resumidamente, programa ABC. A instauração desse programa representou umas das primeiras ações de internalização dos compromissos voluntariamente assumidos pelo Brasil na Conferência das Partes sobre o Clima (COP-15), realizada entre 7 e 8 de dezembro de 2009 em Copenhague na Dinamarca. Entre os compromissos assumidos pelo então presidente do Brasil, Luis Inácio Lula da Silva, estava a redução, até o ano de 2020, entre 36,1 e 38,9% nas emissões dos principais gases causadores do chamado efeito estufa (GEE): dióxido de carbono (CO_2), metano (CH_4) e óxido nitroso (N_2O). Alheio a toda sorte de interpretações sobre as consequências sociais, econômicas e ambientais do compromisso voluntário assumido pelo Brasil, tal responsabilidade colocou o país na rota das potencias emergentes com reconhecido senso de responsabilidade ambiental.

Apesar disso, apenas boas intenções não podem servir de base para a avaliação da consecução das ações derivadas da implantação do programa ABC, e a caracterização científica surge como elemento básico buscando reforçar, ou refutar, as potencialidades advogadas. Dentre as potencialidades que precisam ser estudadas estão aquelas representando benefícios oriundos da substituição de pastagens em diferentes estádios de degradação por agroecossistemas integrando lavoura-pastagem-floresta, sistemas iLPF, um dos pilares do Programa ABC.Nesse contexto, a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) dispõe de unidades de referências tecnológica (URTs) espalhadas por todo o Brasil, onde sistemas iLPF com diferentes designs foram, ou vem sendo, implementados para estudos sobre as consequências sociais, produtivas, e climáticas dessa tipo de agroecossistema integrado. Todo esse esforço vem sendo feito por meio de chamadas de editais de agências públicas de fomento à pesquisa, sejam federais ou estaduais e, dentre deste universo, o projeto intitulado "Agricultura Sustentável no Cerrado e na Transição e na Transição Cerrado-Amazônia" (Projeto Carbioma, código Embrapa 02.11.05.001.00.00 / CNPq 562601/2010-4).

Liderado pela pesquisadora Beata Emoki Madari (Embrapa Arroz e Feijão), o projeto Carbioma tem como objetivos principais: i) descrever como os processos biofísicos do solo são afetados por práticas de manejo do solo, e qual o impacto líquido destas práticas sobre o seqüestro de carbono e as emissões de GEE sob os diferentes usos e manejos da terra; ii) identificar indicadores (biológicos, químicos ou físicos) adequados para a detecção da mudança na qualidade do solo no curto, médio e longo prazo. Particularmente o segundo objetivo é por onde gravitam as ações da corrente tese, que corresponde fazer uso dos dados biológicos, químicos e físicos levantados para explorar e mostrar as potencialidades de uso de abordagem estatística multivariada ainda pouco difundida no âmbito da ecologia de planta e solo: a análise Procrustes.

A tese é apresentada em três capítulos, todos na língua inglesa, uma vez que os textos dos dois primeiros capítulos já foram publicados em língua estrangeira. A parte introdutória e uma breve revisão de literatura foram acrescentadas e redigidas em português. Alterações de formato nos títulos, figuras e no formato de referências bibliográficas referem-se a demandas dos periódicos onde os artigos foram submetidos e publicados. O item de conclusões, em português, sintetiza as principais conclusões e considerações apresentadas nos três capítulos.

O primeiro capítulo descreverá roteiros esquemáticos mostrando como o vetor de resíduos derivado da correlação e duas tabelas de dados pela análise Procrustes (chamado

PAM: *Procrustes association metric*) pode ser utilizado como representante univariado da correlação em outras abordagens estatísticas (ordenação ecológica, regressão, e ANOVA seguida de teste de médias).

O segundo capítulo da tese, utilizando sugestões do primeiro capítulo, tratará de um estudo de caso diretamente ligado ao CARBIOMA. Neste caso, fazenda experimental situada no município de Cachoeira Dourada – GO, e contendo quatro diferentes tipos de uso da terra, dentre os quais um sistema iLPF, foi escolhido para a condução do estudo de caso. O objetivo geral foi acessar como correlações, no formato de PAM, entre tabelas de dados representadas por variáveis individuais de estrutura microbiana (dada por análise de lipídios oriundos do solo; PLFA: *Phospholipids Fatty Acid*) e propriedades individuais de química e física de solo, eram moduladas pelo tipo de uso da terra: pastagem degradada, pastagem melhorada, fragmento de mata nativa, e sistema iLPF. A hipótese para o estudo de caso foi a de que a relação fungo: bactéria, comumente associada a ambientes mais conservativos, é promovida pelo sistema iLPF uma vez que tais sistemas são caracterizados pelo aumento da heterogeneidade vegetal oriunda da sistematizada introdução de espécies arbóreas em meio a pastagem.

O terceiro e último capítulo da tese foi estritamente dedicado a responder questionamentos técnicos referentes à abordagem Procrusteana e surgidos depois das publicações dos dois primeiro capítulos da tese. Neste caso, dois dos questionamentos mais comuns foram abordados. Foram eles: **i**) quais são os efeitos da correlação entre colunas/variáveis dentro de uma tabela de dados sobre os resultados da análise Procrustes? **ii**) Pode o vetor de resíduos procrusteanos, a PAM, traduzir diferenças entre tratamentos em termos da força de correlação multivariada entre duas tabelas de dados?

2. REVISÃO DE LITERATURA

A atividade agrosilvipastoril está entre as principais fontes de carbono antropogênico no Brasil (Lapola et al., 2014), e o país, como potência agrícola e um dos poucos países que ainda dispõem de áreas para expansão da produção de grãos, proteínas, fibras e agroenergia, vem deslocando grande parte de seus esforços para a construção de modelos de agricultura mais compatíveis com as demandas ambientais.

O programa Agricultura de Baixa Emissão de Carbono, Programa ABC, coopta esforços de sustentabilidade agrícola tendo como hastes principais: a fixação biológica de nitrogênio; o plantio direto; a produção e manutenção de florestas; a recuperação de pastagens degradadas; o tratamento de resíduos animais; e a utilização de espécies geneticamente adaptadas às condições locais. Além disso, há o fomento de sistemas de integração entre lavoura, pecuária, e floresta, doravante chamado iLPF. No que diz respeito ao iLPF, esta talvez seja a haste do programa ABC que mais se adeque as aspirações brasileiras para o modelo de agricultura nacional no corrente século. Isso devido ao seu caráter aglutinador ao se tratar de sistema agrícola que combina cultivo arbóreo, de grãos, e criação de animais de forma simultânea e/ou sequencial, com uso intensivo e sustentável da terra, e que proporciona a máxima produção de alimentos, fibras e energia por unidade de área. Segundo Balbinoet al. (2012), a integração entre lavoura, pecuária, e floresta, está inserida no conceito dos Sistemas Agrosilvipastoris e, quando associada à práticas conservacionistas como o plantio direto na palha, tem sido postulada como alternativa potencialmente viável para a recuperação de pastagens degradadas.

Ao redor do mundo, as terras sob o uso de pastagens respondem por cerca de 12% do carbono total existente na biosfera (Schlesinger, 1977), sendo que 90% desse montante estão alocados no solo, seja na biomassa radicular, seja na forma de matéria orgânica associada à rizosfera (Parton et al., 1993).Sendo assim, tais ecossistemas podem compreender importante dreno de carbono atmosférico quando bem manejados (Connant et al., 2001; Lal et al., 2007). Contudo, o modelo de produção animal no Brasil, país com o maior rebanho bovino do planeta, traz o ranço histórico do uso irracional dos recursos naturais e, como fruto desse viés, é estimado que cerca de 50% da área cultivada com pastos no Cerrado e mais de 60% da área de pastagem existente na Amazônia se encontram em algum estádio de degradação (Dias-Filho & Andrade, 2006; Strassburg et al., 2014). Infelizmente, apesar dos significativos avanços obtidos nos últimos anos no combate ao desmatamento, a abertura de áreas para a utilização na pecuária extensiva retrata uma realidade ainda em voga no Brasil, representando fator de pressão sobre importantes biomas como o Cerrado e a Amazônia (Strassburg et al., 2014, Lapola et al., 2014).

A conversão de florestas em pastagens e/ou lavouras é acompanhada da drástica alteração na ecofisionomia da paisagem, sendo a redução da biomassa vegetal o ponto mais dramático (Wallenius et al., 2011). Florestas podem sequestrar substanciais quantidades de carbono em sua biomassa, e a redução na biomassa vegetal como resultado da conversão de florestas em pastagens e/ou lavoura pode conduzir a uma sensível redução nos estoques de C (Guo & Gifford, 2002).Além da redução na biomassa vegetal, a conversão de floresta para pastagem e/ ou lavoura vem associada à redução da diversidade de espécies vegetais, configurando um cenário de maior homogeneidade nas condições físicas e químicas do solo, incluindo a quantidade e qualidade da matéria orgânica aportada. Esse cenário de redução da diversidade heterogeneidade vegetal como resultado da conversão de florestas em pastagens foi argumentado como capaz de causar a “homogeneização” da comunidade microbiana do solo (Rodrigues et al., 2013). De fato, está bem sedimentado o papel da comunidade

microbiana como responsável pela maior parte dos serviços dos ecossistemas relacionados ao ambiente do solo, tendo na matéria orgânica a principal arena para as suas interações com as plantas e outros membros da biota (Wallenius et al., 2011). Por isso, é normalmente hipotetizado que as mudanças no uso da terra, ao afetarem as condições químicas e físicas do solo, afetam de forma sensível o tamanho, a atividade e a composição do espectro microbiano (Fernandes et al., 2011; Vallejo et al., 2012). Assim, devido à rápida resposta às alterações nas condições ambientais, aspectos microbiológicos do solo, como a biomassa e a estrutura da comunidade microbiana são apresentados por diversos autores como potenciais indicadores de mudanças no uso da terra (Acosta-Martínez et al., 2010; Fernandes et al., 2011; Vallejo et al., 2012; Moeskops et al., 2012).

Dentro do diversificado espectro microbiano presente no solo, fungos e bactérias formam a base dos dois principais canais de decomposição e transferência de energia para as redes tróficas da fauna do solo (Scharroba et al., 2012), compreendendo 90% da biomassa microbiana total (Six et al., 2006; Rinnam & Bååth, 2009). Tem sido sugerido que os solos com domínio de fungos são comumente associados a menores taxas de mineralização do carbono do que aqueles com domínio de bactérias (Bardgett & McAllister, 1999; Six et al., 2006; van der Heijden et al., 2008; de Vries et al, 2011). As explicações normalmente recaem sobre as diferenças estequiométricas existentes na biomassa desses dois grupos (van der Heijden et al., 2008). Enquanto a biomassa bacteriana é constituída basicamente de polímeros de fácil degradação, baixa relação C:N; fungos são estruturados com polímeros de baixa labilidade, ou seja, maior relação C:N (Baldrian et al., 2006). Tais diferenças influenciam nas preferências de fungos e bactérias quanto à qualidade dos resíduos orgânicos necessários ao crescimento e manutenção (Rousk & Bååth, 2007) e, por consequência, em suas taxas de mineralização (van der Heijden et al., 2008; De Dyen et al., 2008). Isso reforça a visão difundida de que fungos possuem maior eficiência de crescimento do que bactérias, ou seja, produzem mais biomassa por unidade de carbono mineralizado a partir do substrato orgânico (Six et al., 2006); sendo comum os estudos sugerindo que sistemas de manejo habilidosos em armazenar carbono, são também aqueles com maior dominância relativa de fungos (Bardgett & McAllister, 1999; de Vries & Bardgett, 2012). Estudo conduzidos por de Vries et al. (2011), por exemplo, apontou a menor taxa de mineralização de carbono e nitrogênio em solos dominados por fungos. Consequentemente, o deslocamento da comunidade microbiana em direção ao domínio de fungos ou bactérias poderia antecipar eventos de conservação ou perda de carbono pelo solo, servindo como indicador de qualidade de tipos de uso da terra, como no caso dos sistemas de integração lavoura-pecuária-floresta, cerne do estudo de caso a ser apresentado no segundo capítulo da tese.

Ecossistemas de pastagens são caracterizados pela alta densidade de raízes nos primeiros centímetros do solo, o que contribui para a transferência para o solo de grande quantidade de formas de carbono facilmente decomponíveis (Millard & Singh, 2010). Estudos têm levantado evidências que substanciam a hipótese de que em solos sob influência da pastagemhá o domínio de bactérias em relação aos fungos, atribuindo tal resultado ao maior aporte de carbono lábil em comparação aquele fornecido por vegetação de floresta; além das perturbações resultantes do aporte de nutrientes minerais, da mecanização do solo, e do pastejo (Potthast et al., 2011,2012). Assim, dentro de um sistema que integre pastagem manejada com uma ou mais culturas florestais, como é o caso dos sistemas de iLPF no Brasil, é esperada a variação espacial na dominância relativa de fungos e bactérias, com bactérias prevalecendo no pasto e fungos sendo mais abundantes sob a área de influência do componente arbóreo da integração, a qual recebe menos aporte de nutrientes e não é constantemente perturbada pelas operações de manejo. Essa hipótese foi testada por Vallejo et al. (2012) ao conduzirem investigação sobre comunidade microbiana presente fora e abaixo da área de influência do dossel de *Prosopis juliflora* dentro de uma cronosequência de

sistemas integrando pasto e silvicultura na Colômbia. Apesar de encontrarem resultados que sustentaram a hipótese estabelecida, atribuindo-os ao específico ambiente químico e físico no solo sob a influência arbórea, tais resultados, obviamente não podem ser extrapolados para bases globais. Isso porque dentro do da faixa de opções de espécies arbóreas e/ou arbustivas para a constituição do componente florestal nos modelos de integração, as diferenças ecofisiológicas existentes, como a produção e qualidade de liteira e rizodepósitos, podem gerar padrões distintos na variação na estrutura da comunidade microbiana. No Brasil, por exemplo, o gênero mais comumente utilizado para a composição do extrato florestal em sistemas de integração é o *Eucalyptos*, e as justificativas são muitas, como o grande número de espécies; os múltiplos usos; e a considerável plasticidade ecológica e o rápido crescimento (Balbino et al., 2012). Contudo, outras opções para a formação do componente florestal em sistemas iLPF vêm ganhando força, como o pinho-cuiabano, o paricá, o mogno africano, a acácia, o cedro australiano, a leucaena, e a seringueira. Dessa forma, o contexto brasileiro enseja estudos atentando para a comparação de diferentes tipos de sistemas iLPF, algo que ainda não foi documentado na literatura. Por exemplo, em recente artigo, Salton et al. (2014) levantaram dados derivados de quatorze anos de estudos em sistemas iLPF, os quais suportaram os postulados advogando a superioridade agronômica e ecológica dos sistemas iLPF em relação às pastagens mal manejadas, as quais, em seu turno, são uma tendência no Brasil (Lapola et al., 2014; Strassburg et al., 2014).

A visualização das comunidades microbianas e dos seus aspectos – biomassa e estrutura - pode ser feita a partir de diferentes vitrines, representadas pelas metodologias atualmente disponíveis. Reunindo vantagens e desvantagens, a escolha da metodologia apropriada terá como base fundamental o objetivo do estudo e a familiaridade do pesquisador com a técnica. Entre as opções que mais cresceram para a condução de estudos de ecologia microbiana do solo está a análise de ácidos graxos derivados de lipídios extraídos do solo.Ácidos graxos são basicamente ácidos carboxílicos com longas cadeias hidrocarbonadas, derivados principalmente de moléculas lipídicas. Lipídios, por sua vez, correspondem a um amplo conjunto de biomoléculas com diversas funções nas células vivas, caracterizadas principalmente pela alta solubilidade em solventes orgânicos e pela baixa solubilidade em água. Entre os principais tipos de lipídios existentes nas células vivas estão aqueles associados à estruturação das membranas celulares, os fosfolipídios (PLFA do inglês "Phospholipids Fatty Acid"); e aqueles vinculados ao armazenamento de energia, os chamados lipídios neutros (NLFA do inglês "Neutral Lipids Fatty Acid"). Esses últimos tendo como principais representantes os triacilgliceríd ios, presentes exclusivamente em células eucarióticas (Ruess and Chamberlain, 2010). Ambos os tipos, fosfolipídios (PLFA) e neutrolipídios (NLFA), são fontes de ácidos graxos de uso potencial como marcadores microbianos do solo (Bååth, 2003).

Os fosfolipídios são os componentes estruturais mais importantes das membranas biológicas, sendo responsivos à variação em tamanho das células microbianas ou, equivalente, ao tamanho da biomassa microbiana. Por serem rapidamente degradados após a morte das células (Pinkart et al., 2002), e representarem uma proporção relativamente constante da biomassa de microrganismos, a soma total dos marcadores graxos é comumente utilizada como uma estimativa da biomassa microbiana ativa do solo (Zelles et al., 1999). A estimativa da biomassa microbiana por ácidos graxos de fosfolipídios tem se mostrado altamente correlacionada a biomassa obtida por métodos mais tradicionais (van Groenigen et al., 2010; Langer & Rinklebe, 2011). Com relação à outra fração geradora de marcadores lipídicos (NLFA), são poucos os estudos que a consideram, sendo a maior parte deles conduzida sob condições de laboratório (ver Olsson et al., 1999 e Bääth, 2003). Baseado em estudos de Neidhart et al. (1990), os quais observaram que bactérias não armazenam energia na forma de lipídios neutros, tem sido sugerido o uso da variação na relação NLFA/PLFA como indexador do status fisiológico dos fungos no solo (Bååth 2003).

As propriedades variáveis dos ácidos graxos derivados de lipídios extraídos do solo, as quais permitem o seu uso na discriminação entre grupos microbianos pelos ecologistas de solo, estão intimamente relacionadas às suas respectivas rotas de biossíntese. Dessa forma, grupos de ácidos graxos são relativamente particulares para alguns grupos microbianos (ver Ruess & Chamberlain, 2010). Na literatura, as diferentes designações dos ácidos graxos estão associadas às propriedades discriminantes associadas à rota de biossíntese. No entanto, um padrão básico encontrado é o **A: BωC** (Zaady et al., 2010), onde a letra **A** denota o número de átomos de carbono da cadeia; a letra **B** representa o número de instaurações presentes na cadeia carbônica; a letra grega ômega (**ω)** indica o grupo metil (CH_3) terminal da molécula, e a letra **C** indicada a posição da primeira insaturação contada a partir do grupo metil terminal (ω). A título de exemplo, o marcador para fungos em geral (**18:2ω6**) indica dezoito carbonos formando a cadeia, duas insaturações, onde a primeira insaturação está a seis carbonos de distância do grupamento metil terminal da molécula. Algumas outras notações podem ser encontradas na literatura (Tabela 2), como é o caso dos prefixos **a** e **i**, os quais indicam ramificação **anteiso** e **iso**, respectivamente. Os sufixos **t** e **c**, por exemplo, indicam conformação trans e cis, respectivamente. Outros, como o **α-OH** e o **β-OH,** indicam que as hidroxilas estão, respectivamente, na primeira e segunda posição da cadeia contando a partir do grupo carboxílico. Há também a designação em que um número vem seguido da notação **ME**, indicando a posição do grupamento metil (CH_3) na cadeia hidrocarbonada. Designações com prefixo **cy** indicam a existência de grupamento ciclopropano ao longo cadeia. Curiosamente, apesar de ser uma técnica em crescente uso (Frostergard et al., 2011) há escassez de trabalhos de ecologia microbiana do solo conduzidos no Brasil e explicitamente acessando a alterações na estrutura microbiana do solo por meio da PLFA (Fernandes et al., 2011).

Os dados derivados da análise PLFA são naturalmente analisados sob a ótica multivariada. Por exemplo, é disseminado o uso das técnicas de ordenações, como Análises de Componentes Principais (PCA) e Escalonamento multidimensional não métrico (NMDS), para sumarizar a informação contida nas matrizes de dados da PLFA representando a estrutura da comunidade microbiana, e assim detectar padrões globais de distinção entre preditores categóricos. Também, abordagens multivariadas assimétricas (que atribuem papel de explicativa para uma matriz de dados enquanto outra é considerada matriz resposta, Legendre & Legendre, 2012) como Análise de Redundância (RDA), vêm sendo utilizadas para acessar a resposta linear da tabela de dados PLFA à tabela de dados contendo preditores, por exemplo, propriedades químicas e físicas de solo (Bossio et al., 1998;Lundiquist et al., 1999; Hossain & Sugyiama, 2012; Ramsey et al., 2012). Menos comumente utilizadas, porém, são as abordagens multivariadas visando correlacionar tabelas de dados sem a necessidade de atribuir papéis de matriz resposta e matriz preditora (ou explicativa) para qualquer uma delas, consideradas por Legendre & Legendre (2012) como métodos simétricos. Entre elas está análise Procrustes (Gower, 1971). Com respeito à abordagem Procrusteana, ou simplesmente análise Procrustes, esta possui características que a torna atrativa para estudos de ecologia de solo, como o fornecimento de vetor de resíduos representando o relacionamento multivariado na forma univariada e que pode ser utilizado em outras abordagens estatísticas. Contudo, o baixo uso dessa abordagem em ecologia de solo mostra a necessidade de esclarecimentos sobre as potencialidades dessa abordagem em contexto prático, e nada melhor que uma tese de doutorado para isso. Nos próximos capítulos, incluindo o estudo de caso conduzido em Cachoeira Dourada (Tabela 2), a análise Procrustes será apresentada em riqueza de detalhes de cunho teórico e prático. Por isso, não haverá delonga para revisar tal abordagem na revisão de literatura geral a fim de evitar redundância com o que virá a seguir.

Tabela 1. Alguns dos principais ácidos graxos utilizados como marcadores, suas frações lipídicas de origem e origens biológicas predominantes.

Tipo de ácido	Frequentemente encontrados	Fração lipídica de origem	Origem predominante	Referências
Saturados				
≥20c, e não ramificados	22:0; 24:0	PLFA e NLFA	Plantas	ZELLES (1999) RUESS et al. (2007)
Ramificações iso e anteiso	i e a C14-C18	PLFA	Bact. Gram (+)	ZELLES (1999)
Anel de ciclopropano (cy)	cy17:0 e cy19:0	PLFA	Bact. Gram (-)	ZELLES (1999)
Ramificação metil (Me) no carbono 10	10Me C15-C18	PLFA	Bact. redut. SO_4^{-2} Actnomicetos	DOWLING et al. (1986) KERGER et al.(1986)
Hidroxila substituinte	OH em C10-C18	PLFA	Bact. Gram (-) Actinomicetos	WAKEHAM et al., (2003); LEE et al., (2004); MIRZA et al., (1991)
Monoinsaturados				
Insaturação no C5 (ω5)	16:1ω5	PLFA	FMA e Bact.	OLSSON et al. (1999); SAKAMOTO et al., (2004); ZELLES (1997)
		NLFA	FMA	OLSSON et al. (1999); MADAN et al.(2002)
Insaturação no C7 (ω7)	16:1ω7	PLFA	Amplo Bact.	GUCKERT et al. (1991); ZELLES(1999)
	18:1ω7	PLFA	Bact/ FMA	ZELLES, (1999); OLSSON et al., (1999)
Insaturação no C8 (ω8)	18:1ω8	PLFA	Bact. Metano.	RINGELBERG et al. (1989)
	16:1ω8	PLFA	Bact. Metano.	BODELIER et al. (2009)
Insaturação no C9 (ω9)	18:1ω9	PLFA PLFA PLFA/NLFANLFA PLFA	Fungo geral Bact. Gram (-) Plantas Nematóides	BÅÅTH (2003) ZELLES, (1999) RUESS et al. (2007) CHEN et al. (2001)
	20:1ω9		FMA *Gigaspora*	SAKAMOTO et al.(2004)
Polinsaturados				
1º insaturação no C6 (ω6)	18:2ω6,9	PLFA	Fungo geral e EM	FROSTERGARD & BÅÅTH, (2001); ZELLES (1999)
		NLFA	Animais	
	18:3ω6,9,12	PLFA/NLFA PLFA	Plantas Zigomicetos	RUESS et al. (2000) RUESS et al. (2007)
	20:4ω6,9,12,15	PLFA/NLFA	Amplo animal	WESTHUIZEN et al.(1994) CHEN et al.(2001)

PLFA (ácidos graxos derivados de fosfolipídio); NLFA (ácidos graxos derivados de lipídios neutros); FMA (fungos micorrízicos arbusculares); EM (fungos ectomicorrízicos);Bact. red. SO_4^{-2}(Bactérias redutoras de sulfato).

Tabela 2. Características gerais da unidade de referência tecnológica (URT) selecionada para a condução do estudo de caso da presente tese.

Características	**URT - Cachoeira Dourada**
Localização	Cachoeira Dourada - GO
Tipo de iLPF	Agrosilvipastoril
Ano de implantação	2008
Bioma	Cerrado
Tipo de Solo	Latossolo Vermelho
Clima (Köppen)	Aw
Espécies agrícolas	Milho (*Zea mays L*); Soja (*Glycine max* L)
Espécies forrageiras	*Brachiaria brizantha* cv Marandú, Piatã
Espécies florestais	*Eucalyptus urograndis*
Bovino	Nelore, mestiços
Atividades	Carnes; grãos. Mourões, etc

3. CAPÍTULO I:

MUCH BEYOND MANTEL: BRINGING PROCRUSTES ANALYSES TO THE PLANT AND SOIL ECOLOGIST'S TOOLBOX

Capítulo publicado como: Lisboa, F. J. G.; Peres-Neto, P. R.; Chaer, G. M.; Jesus, E. C.; Mitchell, R. J.; Chapman, S. J.; Berbara, R. L. L. Much beyond mantel: bringing Procrustes association metric to the plant and soil ecologist's toolbox. **PLoS One**, v. 9, e101238, 2014

3.1. RESUMO

A correlação de dados multivariados é uma tarefa comum em investigações no âmbito da biologia do solo e da ecologia de maneira geral. A análise Procrustes e o teste de Mantel são duas abordagens que comumente atendem a este objetivo, sendo consideradas análogas em muitas situações, especialmente quando usadas como uma forma de testar a significância estatística da correlação entre duas tabelas de dados multivariados. Aqui, nós chamamos a atenção para as vantagens de aplicação de uma das características da análise Procrustes pobremente explorada: os resíduos Procrusteanos. Esses resíduos representam diferenças entre duas tabelas de dados considerando observações homólogas (por exemplo, pontos amostrais), e podem ser explorados no sentido de estimar níveis de associação individuais (por exemplo, se alguns grupos de amostras são mais similares do que os outros em termos de associação entre tabela de dados microbiológicos e tabela de dados ambientais). Neste artigo, usando dados reais e hipotéticos, busca-se familiarizar ecologistas com os benefícios de usar as diferencias locais em termos de resíduos procrusteanos no sentido de ganhar indícios sobre processos que regulam a associação entre dados multivariados. Uma vez que a informação multivariada é transformada em matriz de dissimilaridade/similaridade quando usando o teste de Mantel, esse teste não permite que os ecologistas contrastem pontos amostrais homólogos ao longo de tabelas multivariadas de interesse, não permitindo a análise de correlação em subsequentes abordagens estatísticas clássicas, como ordenação, particionamento da variância (regressão), e ANOVA.

Palavras-chave: Matriz de dados. Tabela de dados. Ordenação. Análise multivariada. Microbiologia ambiental. Variação residual.

3.2. ABSTRACT

The correlation of multivariate data is a common task in investigations of soil biology and in ecology in general. Procrustes analysis and the Mantel test are two approaches that often meet this objective and are considered analogous in many situations especially when used as a statistical test to assess the statistical significance between multivariate data tables. Here we call attention to ecologists of the advantages of a less familiar application of the Procrustean framework, namely Procrustean residuals. These residuals represent differences in fit between multivariate data tables regarding homologous observations (e.g., sampling sites) that can be used to estimate local levels of association (e.g., some groups of sites are more similar in their association between biotic and environmental features than other groups of sites). In this paper, we attempt to familiarize ecologists with the benefits of using these local residual differences to further gain insights about the processes underlying the association among multivariate data tables using real and hypothetical examples. Given that in the Mantel framework, multivariate information is translated into a pairwise distance matrix, we lose the ability to contrast homologous data points across dimensions and data matrices after their fit.

Key-words: Data matrix. Data tables. Ordination. Multivariate analyses. Environmental microbiology. Residual variation.

3.3. INTRODUCTION

In multidimensional data analysis, ecologists often encounter situations where they need to choose between two or more numerical approaches that are able to tackle the same question of interest. The preference between approaches is based, among other factors, on the familiarity of the user with the method, which in turn depends on the time a particular method has been available in statistical packages and the ease in implementing and interpreting its results. Another relevant factor to consider is "literature–induced use" in which renowned research groups involved in the development, improvement and generation of statistical ecological approaches have a strong influence on the types of statistical approaches other ecologists use.

Determining the strength of the relationships between multivariate datasets is a routine analysis when trying to understand the environmental factors driving the composition and structure of ecological communities. Two approaches, the Mantel test (Mantel, 1969) and Procrustes analysis (Gower, 1971), though considered analogous by the literature in the questions they can tackle (Peres-Neto & Jackson, 2001), have not been used to the same extent. Despite the advantages of Procrustes analysis over the Mantel test (Peres-Neto & Jackson, 2001) regarding greater statistical power in detecting significant relationships (i.e., lower type II errors) and the possibility of analyzing further the patterns of association between multivariate matrices (visually and by further statistical analyses), the Procrustean approach remains relatively unused in tackling questions regarding the relationships between data matrices involving plant and soil information or between soil matrices (Fig. 1).

The Mantel test and the Procrustes approach can be both used in many similar situations where the interest is into assessing how multivariate data matrices are associated (correlated), though for unknown reasons they have been used in quite different ways in the ecological literature. For example, while the Mantel test has been often applied when testing the relationship between above and below ground data matrices (Tuomisto et al. 2003a; Tuomisto et al., 2003b; Tuomisto et al., 2003c; Kang & Mills, 2004;Poulsen et al., 2006; Fitzsimons et al., 2008; Powers et al., 2009; Castilho-Monroy et al., 2011; Pomara et al., 2012), Procrustes analysis has predominantly been used to contrast the results of different ecological ordinations on the same data (Artz et al., 2006; Trivedi et al., 2008; Jesus et al., 2009; Merillä et al., 20010), to compare fingerprinting tools for assessing microbial communities (Grayston et al., 2004; Singh et al., 2006; Vinten et al., 2011) and for deciding between methodological choices (Hirst & Jackson, 2007; Poos & Jackson, 2012). Indeed the Procrustean framework has been rarely used to make inferences about plant and soil relationships (Singh et al., 2008; Burk & Irwin, 2009; Lisboa et al., 2012; Landeiro et al., 2012) and other types of ecological associations between data sets. There are certainly applications in which the Procrustean and Mantel test cannot be easily applied interchangeably. Unlike Mantel, the Procrustean approach can be used when comparing multiple data matrices jointly. Conversely, ecologists are often interested in correlating distance (or similarity) matrices rather than testing the association among data matrices in their raw form (i.e., not transformed by the property of distance measures). One particular case in the distance-decay of similarity in ecological communities (Nekola & White, 1999) in which one is interested in testing the hypothesis that the similarity in community composition decreases in relation to linear (or log transformed) geographic distance between communities. The differences between raw-based and distance-based approaches have been discussed extensively elsewhere (Legendre et al., 2005; Tuomisto & Ruokolainen, 2006).

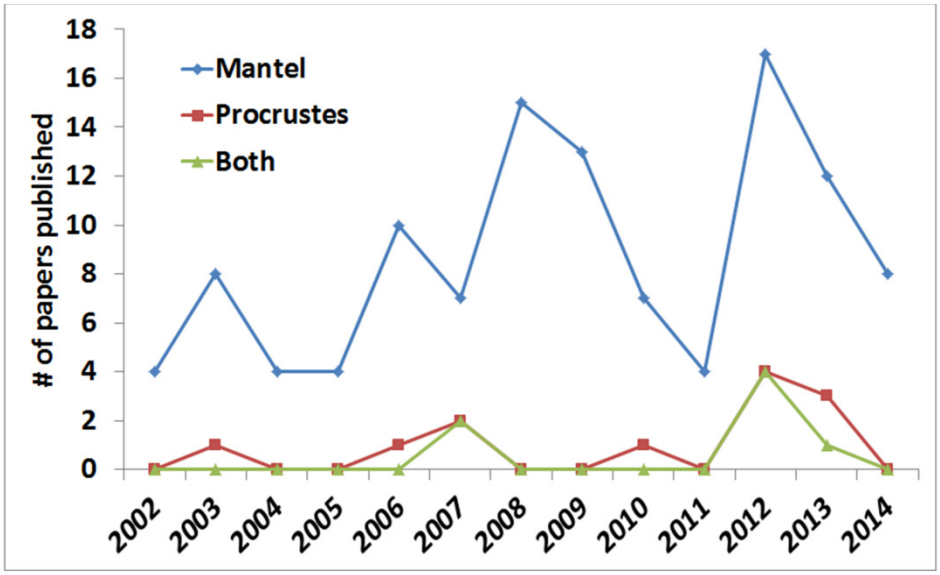


Figure 1. Number (#) of papers published using Mantel and Procrustes for relating data matrices from soil or plant studies in the ten years since Peres-Neto & Jackson (2001) stated the advantages of Procrustes over the Mantel approach. Data obtained using Thompson Reuters database (June, 6, 2013). We searched for papers using uniquely the Mantel approach, uniquely the Procrustes approach and papers using both approaches. The search was based on Procrust* (Procrustean or Procrustes) and Protest.

Despite the relative merits of the Procrustean framework over the Mantel test shown by the relatively well-cited paper by Peres-Neto & Jackson (2001), its potential has not yet been tapped. Perhaps the reason for not getting Procrustes analysis to be as popular as the Mantel test among ecologists yet is the lack of a paper showing that in many situations traditionally investigated by Mantel, the Procrustean analysis can be equally well used. Here, we attempt to familiarize ecologists with the use of Procrustes analysis by using real and hypothetical examples where the Mantel test tends to be preferred. Most importantly, we highlight little explored limits of Procrustes by using its residual vector of association between data tables, hereafter referred as to PAM, in three common statistical approaches: multivariate ordination, variation partitioning and ANOVA.

The use of PAM has been quite restricted in the ecological literature. To our knowledge the first study was by Alárcon et al. (2008) who assessed the plant-pollinator interaction during three consecutive summers in the southeastern portion of California, USA. These authors employed the residual vector (PAM) to identify which pollinating species exhibited the greatest deviation between two consecutive years. Singh et al. (2008) used the PAM in a study on soil microbiology to verify the effect of soil pH on the relationship between arbuscular mycorrhizal fungi (AMF) and plant assemblages. These authors employed the following strategy: 1) Procrustes analysis was applied between the matrices representing the AMF community and that representing the plant community; 2) after detecting a significant relationship (m_{12} = 0,28; P < 0,001); 3) then, these authors extracted the PAM and used it as a response in a simple regression analysis with the soil pH. It was detected no effect of pH on the association between the AMF and plant communities, suggesting that neither the environmental nor the identity of the plant species that composed the community affected the AMF community. Other applications can be certainly found (e.g., Lisboa et al., 2012, Landeiro et al., 2012; Siqueira et al., 2012) but its flexibility and general usage remains largely unexplored.

3.3.1. Procrustes analysis: a foundation for soil and plant ecologists

In ancient Greek mythology there was a character named Procrustes who was a resident of Eleusis Mountain, a known travelers' route. As a "good" host, Procrustes always invited travelers to spend the night at his home; more specifically, he invited them to lie down on his iron bed, which was tailored to fit Procrustes' own body. The guests who did not fit the dimensions of his bed either had their limbs cut off or were stretched until their dimensions approached those of Procrustes's bed. Ironically, none of the guests ever fitted the iron bed because Procrustes secretly had two beds of different sizes (Kuehnelt-Leddihn, 2007). One can easily make a parallel here with ecological data in which data from different sources will almost never easily compare or fit to one another.

Procrustes analysis is based on the search for the best fit between two data tables, hereafter referred to as matrices, where one is kept fixed ("Procrustes' bed" or target matrix), while the other ("Procrustes' guest" or rotated matrix) undergoes a series of transformations (translation, mirror reflection and rotation; Gower , 1971) to fit the fixed matrix. Although in this paper we concentrate on fitting two matrices, the extension of Procrustes analysis to multiple species is straightforward (Peres-Neto & Jackson, 2001) in which the reference matrix can be either one of the original matrices or their averages (or medians). Hereafter, the target matrix (target) will be referred as to **X**, and the data matrix to be fitted as **Y**. **X** and **Y** are both n x p matrices, where n is the number ofrows and p is the number of columns. The goal of the transformations in **Y** is to minimize the residual sum of squared differences between the corresponding n dimensions between **X** and **Y**; the sum of the squares of these residual differences is termed m^2(Gower's statistic), representing the optimal fit between the two data matrices, such that the higher the value of m^2, the weaker the relationship between the two data tables is. The significance of m^2 can be estimated through a permutation test (termed Protest after Jackson (2001); see Peres-Neto &Jackson(2001) for further details).

3.3.2. Procrustean association metric (PAM)

The least squares superimposition between the corresponding n observations of **X** and **Y** is one of the main advantages (in addition to the increased statistical power) of the Procrustean framework in contrast to the Mantel test. The Procrustes superimposition generates a (n x p) matrix of residuals that can be further used to contrast the differences between homologous observations (rows) across matrices in the form of a vector (PAM). Given that within the Mantel approach differences between observations across all dimensions are packed down into a single distances, it cannot be used to assess differences across observations across dimensions. Consistent small and large differences across homologous observations across matrices in regard to other factors of interest can further assist in understanding how **X** and **Y** are related. For example, we could use PAM to assess the degree of observations matching between a plant function trait matrix and a composition matrix and assess whether smaller or greater residuals values are a function of the time elapsed since some disturbance event.

PAM is simply a vector of residual differences between the corresponding n observations. For example, assuming that an ecologist wants to correlate two matrices of data **X** and **Y**, both of which are formed by four rows (i.e. sites, plots, observational units), Procrustes analysis will generate four residual differences between the **X** and **Y** configurations. The compilation of these residual differences between homologous rows (observations) across dimensions in the form of a vector – PAM – represents a useful way to represent information on the relationship between two matrices and make it available for further statistical analysis, both parametric and non-parametric; this feature is not offered by the Mantel approach.

3.4. ROADMAP FOR APPLYING PAM

3.4.1. Constructing a practical roadmap for applying PAM

There are few studies in the ecological literature that have used PAM for analyzing relationships between plant and soil datasets. The lack of examples partially explains the low popularity of Procrustes analysis among plant and soil ecologists and ecologists in general as an alternative tool to the more traditional Mantel test. In order to make the possible uses of Procrusteanresiduals more familiar, we will introduce a number of examples in the form of schematic roadmaps for applying PAM in association with three common statistical approaches: ordination, regression analysis and ANOVA. All codes for the following roadmap can be found out at:

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0101238

Plant and soil ecologists must keep in mind that Procrustes analysis requires that the **X** and **Y** have the same number of rows and columns, though the last dimension is less restricting (see below). Given that the data for both matrices usually originate from the same sites, it is most common in ecology that only the number of columns (descriptors or variables) varies between the two matrices. Therefore, the question arises of how to make the number of columns equal across the two matrices, i.e., how to reduce them to the same dimensionality. Although Procrustes analysis can be performed between matrices having different number of dimensions (i.e., the fit is based on a singular value decomposition (svd) of $\mathbf{X}^{\mathbf{T}}\mathbf{Y}$, where **X** and **Y** are scaled prior to svd and **T** stands for matrix transpose), traditionally the matrix with less number of columns ("missing columns") is made equal in dimension to the larger matrix by adding columns of zeros in order to keep (Fig. 2a; Gower, 1971). Although there are some criticisms related to this practice and alternatives have been suggested (ten Berge et al.,1993), the addition of zero columns does not affect the distances between columns among observations and is a convenient device rather than a hurdle (Dijksterhuis & Gower, 1992).

Another convenient way to make **X** and **Y** have the same number of columns is to represent most of the variation in their raw data by matrices formed by the same number of orthogonal axes (Fig. 2b; Peres-Neto & Jackson, 2001; ten Berge et al.,1993 Dijksterhuis & Gower, 1992), i.e., matrices formed by axes derived through ordination methods such as Principal components analysis (PCA), Non-metric Dimensional Scaling (NMDS), Correspondence Analysis (CA), Principal Coordinate Analysis (PCoA), the choice being dependent on the nature of the data (continuous, presence-absence data, abundance data). Moreover, raw data matrices can previously be transformed into ordination axesin order tomatch data characteristics(see Legendre & Gallagher, 2001); or alternatively have pairwise distance matrices calculated from the data matrices that are then orthogonolized via PCoA to extract ordination axes based on the chosen distance measure (e.g., Bray-Curtis, Jaccard, Sorensen, Gower).

Here, for simplicity, we use a PCA in all applications. In cases, where species data (presence/absence or abundance) was used, the data was Hellinger-transformed and PCAs were extracted on species correlation matrix calculated from the transformed data. The Hellinger transformation alleviates the issue of double-zeros in species data matrix transformed into correlation or Euclidean-distance pairwise matrices prior to PCA in which sites sharing no species in common can be found to be more similar than sites sharing a reduced number of species in common (e.g., the horse shoe effect in ordination plots)(see Legendre & Gallagher, 2001).

The general strategy is as follows:

a) Subject the raw data matrices to an ordination method (here PCA but see above for other strategies);

b) After ordinating **X** and **Y**, use the same number of ordination axes for both matrices (Fig. 2b).

Given that the higher the number of ordination axes used, the higher is the amount of variation explained in **X** and **Y**, it would be interesting run the Procrustean analysis sequentially using matrices made up of an increasing number of ordination axes. It could help ecologists check the consistency of the relationship between **X** and **Y** based on different numbers of ordination axes, which will give more reliability to the results.

3.4.2.The use of PAM in ecological ordination

The first form of PAM shown here is based on ordination methods. Ordination is the graphical representation of the variation of objects (sites), descriptors (species/environmental parameters) or both, in a reduced space formed by orthogonal axes (Legendre & Legendre, 2012).

Three matrices for the soil microbial community were obtained for each site: one based on the fatty acid profile of the soil (PLFA analysis), and the other two on the T-RFLP analysis of the communities of fungi and bacteria, respectively. The matrixrepresenting the soil chemistry

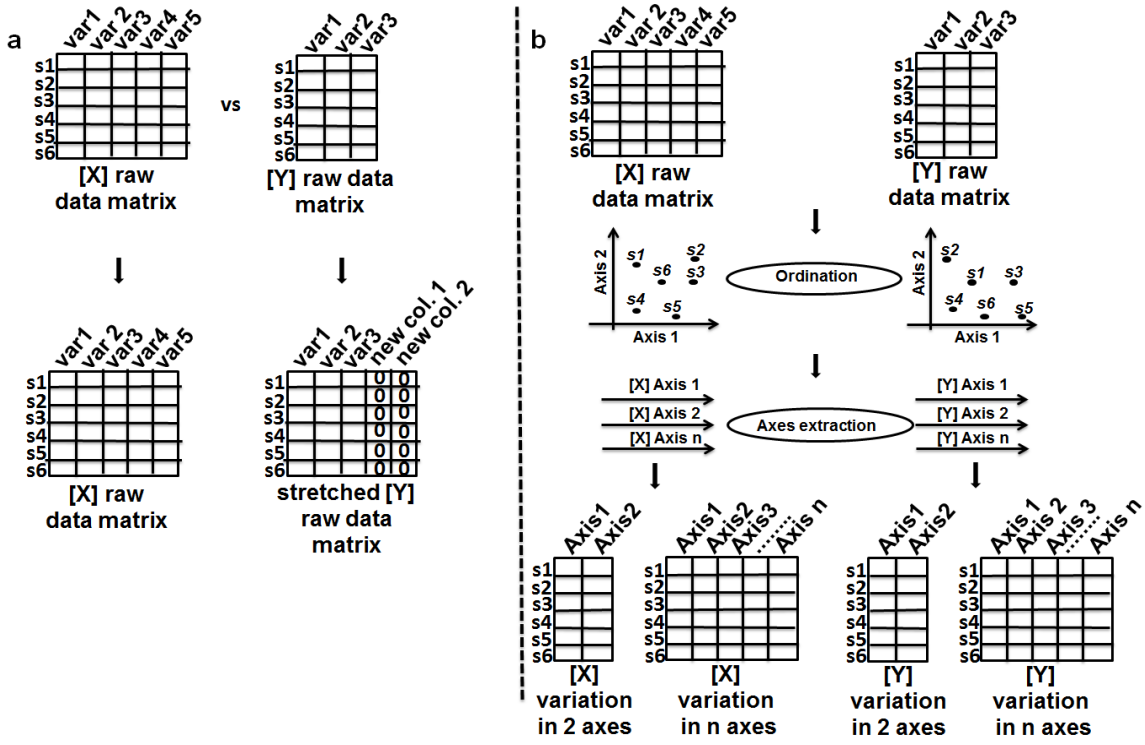


Figure 2.Roadmap for twoalternative waysto reach the same dimensionality between matrices, and so relating it by Procrustes analysis. a) Addition of columns containing zeros to the **Y** raw data matrix for matching the **X** raw data matrix dimension; **b)** Application of ordination to raw data matrices to make matrices have equal dimensionality prior to Procrustes analysis. In the rows, ***s***may stand for sites, plots, sampling points, etc. In the columns, **var** mean variable.

was based on the concentrations of Na, K, Ca, Mg, Fe, Al, P, total C, total N in addition to pH, loss on ignition and moisture.

There is some consensus that the variation in vegetation can act as a proxy for changes in the soil microbial community, either directly in the case of symbionts, for example, or indirectly via changes in soil chemistry itself. We use Procrustes analysis associated with ordination techniques to verify potential drivers of the soil microbial community and to determine if plant community and soil chemistry are equally related to the microbiological variation. The sequence of analysis was as follows:

a) Ordination analysis: All data matrices (community plant, soil chemistry and soil microbial communities) containing the three chronosequences were subjected to separate PCAs based on correlation matrix. The community plant was Hellinger-transformed prior PCA. Then, the first six PCA ordination axes from each matrix were retained in order to assemble four PCA matrices representing the variation summarized in the first 3, 4, 5 and 6 PCA axes. Thus, four PCA matrices were obtained from each dataset: plant community, soil chemistry and soil microbial community (PLFA, bacterial and fungal T-RFLP) (Fig. 3a).
b) Procrustes analysis: The PCA matrices of plant community and soil chemistry were used to run Procrustean analyses with the PCA matrices of soil microbial community based on PLFA, and fungal and bacterial T-RFLP datasets.
c) PAM extraction: Since all Procrustean relationships based on PCA matrices with *n* axes were significant, for simplicity, only the PAM obtained from relationships of PCA matrices with 6 axes were used for subsequent analyses. Six PAMs were generated: PAM1 (soil chemistry on PLFA), PAM2 (soil chemistry on bacteria), PAM3 (soil chemistry on fungi), PAM4 (plant on PLFA), PAM5 (plant on bacteria), and PAM6 (plant on fungi) (Fig. 3c).
d) PAM ordination: The PAMs were assembled in a single matrix ("effect matrix") with one PAM per row (Fig. 3c). Therefore, the effect matrix compiled the effects of plant community and soil chemistry on soil microbial community structure derived from the three methods. This effect matrix was submitted to PCA ordination to verify whether the plant community effect on soil microbial community structure differed from the effect of soil chemistry (Fig. 3c).

The results showed that for all chronosequences the plant effect on microbial structure was divergent in relation to the soil chemistry effect, as suggested by the separation along the axis of greatest variation (Fig. 4). Although we cannot apply a proper statistical significance test in one-table based ordination methods (PCA, NMDS, PCoA, etc), visual inferences can be made. For example the Craggan area exhibited a clear distortion between plant community and soil chemistry variation in terms of their effects on the soil microbial community structure depicted by PLFA, bacterial T-RFLP and fungal T-RFLP (Fig. 4a). Also in this area, the response of the microbial community based on PLFA was distant from the response based on molecular data (T-RFLP) (Fig. 4a).They also suggest that the effects of soil chemical properties on the microbial communities may be weakly mediated by above ground alterations (Lisboa et al., 2012). This example shows the usefulness of Procrustes analysis to raise additional evidence in plant and soil ecology studies.

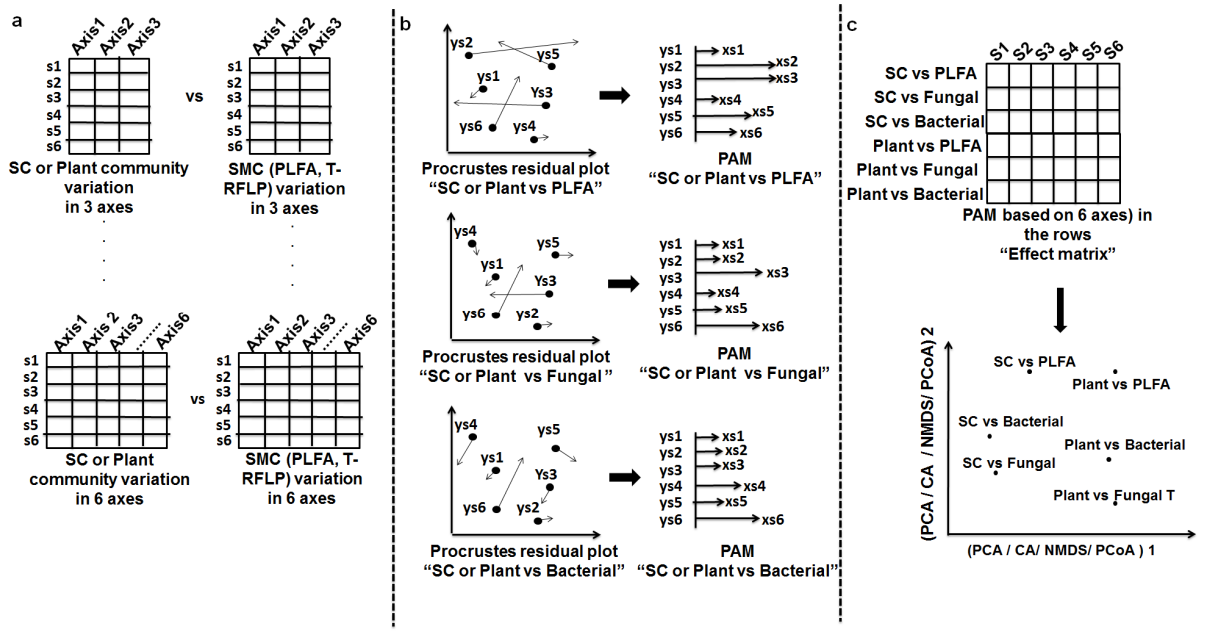


Figure3. Roadmap for applying the Procrustes association metric (PAM) in the multivariate ordination context using data of Mitchell et al. (2010).**a)** Assembling matrices with different ordination axes, through Procrustes analysis, soil chemistry (SC) and plant community with soil microbial community (SMC: PLFA, and bacterial and fungal T-RFLP); **b)**Extraction of PAM from Procrustean relationships based on matrices with 6 ordination axes; **c)** Assembling of PAM based PCA matrices with 6 axes as rows in a single matrix ("effect matrix"), and using it in an ordination technique (e.g., PCA, PCoA, NMDS) to verify if the different effects diverge. In the matrix rows ***s***may stand for sites, plots, sampling points, etc. In the Procrustes residual plot the**ys'**sarethe***s*** scores for the ordination of the **Y**matrixand are pointing towards scores for the ordination of the **X** matrix**.**

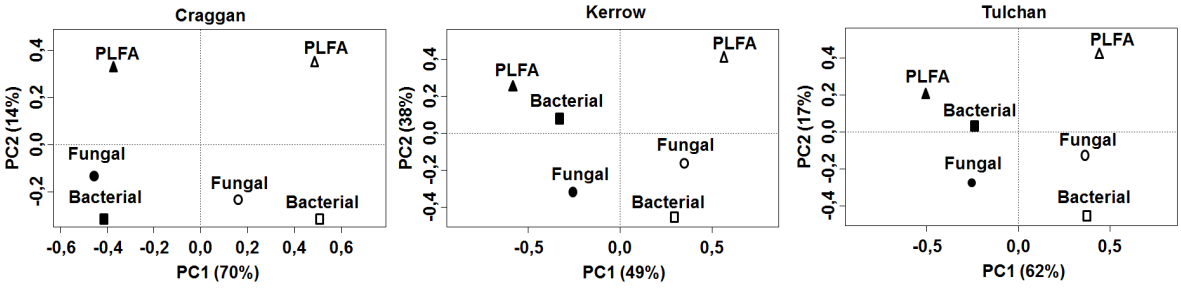


Figure 4. Results from PCA ordination of the Procrustes association metric matrix ("effect matrix") gathering the interactions of soil chemistry and plant community with soil microbial matrices (PLFA, and bacterial and fungal T-RFLP). The **filled symbols** are the Procrustes relationships between soil chemistry and soil microbial matrices, and the **open symbols** between plant community and soil microbial matrices. Data from three chronosequences (Craggan, Kerrow and Tulchan) obtained by Mitchell et al. (2010).

3.4.3. The PAM and regression analysis

In regression analysis, 'response' and 'predictor' are common terms. In ecology, predictors can have different natures. Space, time, organic matter and moisture, among other factors, are some examples of predictors. On the other hand the microbial communities are often used as a response variable because they are considered better indicators of a given ecosystem.

Some authors familiar with soil microbial ecology have been using the Mantel test to assess the individual contribution of deterministic and stochastic processes on the soil microbial structure variation (Dumbrell et al., 2009; Zheng et al., 2013). As an example of the utility of the Procrustes analysis in the context of variation partitioning we can take a hypothetical scenario with four datasets from a given area, corresponding to soil microbial community structure (PLFA), soil microbial functioning (enzyme activities), soil properties and spatial variation. Spatial variation can be represented, for example, by 100 sampling points generated from a 10 m x 10 m transects. The matrix of geographical coordinates of the sampling points can be submitted to PCNM (principal coordinate neighbour matrix) analysis generating a matrix of spatial eigenfunctions termed PCNMs (ten Berge et al., 1993). In this scenario, we can assume that the ecologist aims to assess the relative contributions of individual soil properties (deterministic processes) and spatial variation (stochastic event) on the relationship between microbial community structure and soil microbial functioning rather than on these components individually. To use the Procrustean association metric (PAM) in this context, one can use the following steps:

a) Ordinate the two matrices (i.e., the soil microbial community and soil microbial functioning) via PCA (the soil microbial community matrix was Hellinger-Transformed) and select a similar number of ordination axes (Fig. 5a). The multivariate scores of the two matrices across the selected number of axes are subjected to a Procrustes analysis and a PAM was then calculated.
b) Use individual PAMs (based on 2, 3 or more PCA axes) as response variable and soil properties and spatial variation as independent (predictor) variables in a multiple regression framework (Fig. 5b).
c) Finally, the independent contributions of soil properties (independent of space) and unmeasured spatial process and/or factors (spatial variation independent of soil properties) to the microbial structure can be estimated via variation partitioning (Peres-Neto et al., 2006) and represented by a Venn diagram (Fig. 5c).

3.4.4. The PAM and analysis of variance

Although regression and analysis of variance are ultimately the same analysis in which the response is either continuous (regression) or ascribed to factors (ANOVA), we provide examples for each of them in different sections given that often they are seen as distinct forms of analyses. Evaluation of the effects of land use on soil microbial communities has been a common case-study issue in soil ecology. Some of these studies have been carried out using Mantel approach (Chaer et al., 2008; Peixoto et al., 2010) for doing inferences about land use type effects on soil microbial structure e functioning. However , Mantel does not yield a vector of structure – functioning relationship, that is, a continuous variable, able to be partitioned by categorical variables like land use types. In the following example we show how to use PAM to evaluate the effect of land use type on the relationship between microbial community structure and microbial function in the form of PAM.

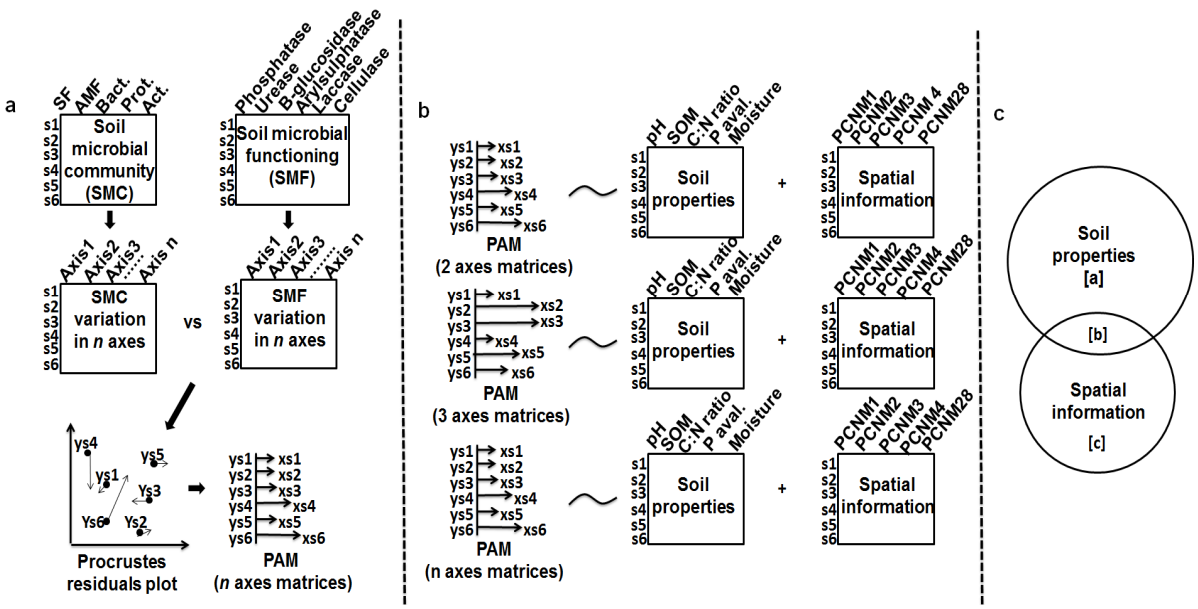


Figure 5.Roadmap forusing Procrustes Association Metric (PAM) in a multiple regression analysis framework (variation partitioning). **a)** Soil microbial community (SMC) and soil microbial functioning (SMF) matrices are submitted to an ordination to reach same dimensionality, and SMC and SMF matrices formed by 2, 3 and *n* axes related through Procrustes analysis in order to generate PAMs; **b)** PAMs generated were used as response in a variation partitioning to verify the individual contribution of soil properties and spatial information (PCNM eigenfunctions) on the SMC-SMF relationship; **c)** Venn diagram depicting the relative contribution of soil properties (niche processes **[a])** and unmeasured spatial factors (neutral processes **[c]).[b]** is the variation in the response due to join contribution of deterministic and neutral processes**.** In the matrix rows ***s***may stand for sites, plots, sampling points, etc. In the Procrustes residual plot the **ys's** are the ***s*** scores for the ordination of the **Y**matrixand are pointing towards scores for the ordination of the **X** matrix**.**

In a hypothetical scenario, a researcher is interested in studying whether four different land use types within the Amazon biome are affecting the relationship between microbial structure and microbial functioning. In each of the land uses (original forest fragment, silvipastoral system, improved pasture, and unimproved pasture) six plots (10 m x 10 m) were established and one composite soil sample (0-10 cm) collected per plot (Fig. 6a). The X dataset (soil microbial structure) was represented by PLFA data, and the Y dataset (microbialfunctioning) by the abundance of genes associated with microorganisms involved in greenhouse gas emission processes, such as nitrifiers, denitrifiers and methanotrophic organisms. The researcher's hypothesis is that in the forest fragment (non-altered environment) there is a better matching between microbial structure and microbial function. Thus, in anthropogenically disturbed environments (silvipastoral system, improved pasture, and unimproved pasture) the change in microbial structure relative to the original (forest) is not followed by a change in the microbial functioning to the same magnitude. This hypothesis can be tested using an integration of Procrustes analysis and ANOVA through the following steps:

a) Reduce the datasets **X** (soil microbial structure) and **Y** (soil microbial functioning) to similar dimensions using PCA. Then, run the Procrustean analysis between the PCA

matrix of the soil microbial community structure and the PCA matrix of soil microbial functioning and extract the PAM (Fig. 6a).

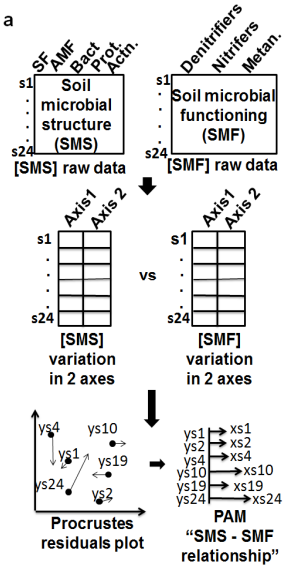
b) Run an ANOVA with land use type as fixed factor and the PAM as the response variable (Fig. 6b).

c) If the F value of ANOVA is significant, a means test can be performed to compare the mean PAMs of the land use types (Fig. 6c).

3.5. DISCUSSION

In this paper we have attempted to show the advantages of the Procrustean analysis over the Mantel test, in which the former can be used for gaining further information on underlying drivers of data table associations. We concentrated on the advantage of that patterns of concordance between data matrices can be displayed and individual observations contrasted separately using the Procrustean framework, allowing further examination of the common and different association patterns among multiple data matrices. Given that in the Mantel framework, multivariate information is translated into a pairwise distance matrix, welose the ability to contrast homologous data points across dimensions and data matrices. It is important to notice that it was not our goal to show the statistical advantages of Procrustes over Mantel as done by previous work (Peres-Neto & Jackson, 2001). Instead, we concentrated on generating different analytical schemes, especially for plant and soil ecologists, to incorporate Procustes in their statistical toolbox.

a



b

Table. One-way ANOVA of the effect of different land use types on the Procrustean relationship between soil microbial structure (SMS) and soil microbial functioning (SMF).

Land use type (fixed factor)	PAM (SMS-SMF relationship)	
	F	*P*
Land use	7.047	0.003**
Residuals	1.345	>0.05

c

Table. Mean effect of different land use types on the Procrustean relationship between soil microbial structure (SMS) and soil microbial functioning (SMF).

Land use types	SMS-SMF relation (PAM in 2 axes)	
	Mean	SE
Original forest	0.021 a	0.003
Silvopastoral system	0.052 b	0.001
Improved pasture	0.074 c	0.003
Unimproved pasture	0.082 c	0.002

Figure 6.Roadmap for using Procrustes association metric (PAM) in an ANOVA context. **a)** PCA ordination of each SMC and SMF raw data matrices, and then Procrustes correlation from 2 axes-based PCA matrices in order to generate the PAM depicting the SMC-SMF relationship.**b**) Table showing results of a One-way ANOVA for using PAM as response and land use type as fixed factor. **c)** Multiple comparisons test (Tukey, 95%) for means of the Procrustean relationship between soil microbial structure and functioning (PAM in 2 axes) across land use types.

What is unique about Procrustean framework? There are at least four characteristics of the approach not shared by others. First, because the approach is correlative rather than

regressive, the number of observations (e.g., sites) in the matrices does not have to be greater than the number of columns as in common regression approaches such as RDA and CCA. Second, we can fit as many matrices as we have available; this latter issue is particularly restrictive under a regression approach given the limitation of number of rows versus number of columns. Moreover, all matrices are treated in equal footing as no matrix is treated as response or predictor. Third, the relationships within (only across) matrix columns do not affect the analysis. Fourth, residual values across observations and dimensions can be calculated as explored as shown here. These characteristics should not be necessarily seen as advantages per se over other methods but rather features that are unique and may be useful in many situations. There are certainly other tools that can be used to look at the associations between data sets. RDA and CCA are well-established tools in ecology and are based on regression (asymmetric) methods. It has been long-standing these approaches may be more appropriate for application to the examples given in this paper, since they establish relations of cause and effect. However, because these analyses include a regression step, they are limited to situations where the number of rows (sites) in the environmental matrix **X** is higher than the number of columns (variables) (Dray et al., 2006; Thioulouse et al, 2004). This is not a limitation in Procrustes analysis and moreover, it is not clear how residual variation among homologous observations across dimensions should be explored in the case of RDA and CCA.

At least two other symmetric approaches are similar to the Procrustean approach, namely Co-inertia analysis (Dolédec & Chessel, 1994), and symmetric Co-correspondence analysis (Ter Braak & Schaffers, 2004) a form of Co-inertia analysis in which a correspondence analysis is applied to two species matrices prior to the analysis. The main difference resides in the fact that fit is influenced by all variables pairs in Co-inertia analysis (within and between matrices), whereas fit is influenced only by variation between matrices in Procrustean. Co-inertia is always based on ordination within data matrices, whereas in Procrustes either the raw data or their ordination axes can be used. Co-inertia can also take into account row (e.g., sites) and column (e.g., species) weights in the analysis, though the standardization and fit processes in Procrustean analysis could also take these into account (Dray et al., 2003). Co-inertia and Procrustean analysis are certainly related in the sense that they both treat matrices as symmetrically during the fitting process, though more studies are necessary to assess which conditions (e.g., correlation within and across matrices, differences in dimensionality between matrices, outliers within and across matrices) they differ. Finally Dray et al. (2003) showed the advantages or merging Co-inertia and Procrustean analysis, whereas the latter is used as a precursor of the former. In reality, future studies are required to contrast the Co-inertia and Procrustean analysis, but in either form of analyses we can produce residual vectors (PAM) that can be further analyzed.

Procrustes can be perhaps best justified when the number of predictors is greater than the number of observations or when **X** and **Y** matrices are equally applicable as explanatory and response variables. In plant-soil ecology, for example, above- and below-ground data matrices can be interchanged as explanatory and response variables. Plant community variation has been shown to be related to variation in below-ground compartments (Lisboa et al., 2012). In addition, soil components such as fertility and the microbial community have been proven to influence aspects of vegetation (van der Heijden et al., 1998). Thus, with the literature showing that both types of datasets under analysis can structure each other, the use of Procrustes analysis, as a symmetric canonical analysis method, should be encouraged among plant and soil ecologists and ecologists in general. We hope that this paper have provided enough examples of the potential for using the Procrustes framework as a precursor to further explore ecological data.

3.6. REFERENCES

ALÁRCON, R.; WASER, N. M.; ORLLETON, J. Year-to-year variation in the topology of a plant - pollinator interaction network. **Oikos**, v. 117, p. 1796-1807, 2008

ARTZ, R. R. E.; CHAPMAN, S. J.; CAMPBELL, C. D. Substrate utilization profiles of microbial communities in peat are depth dependent and correlate with whole soil FTIR profiles. **Soil Biology Biochemistry**, v. 38, p. 2958-2962, 2006.

BURKE, L.; IRWIN, R. The importance of interannual variation and bottom up nitrogen enrichment for plant – pollinator networks.**Oikos**, v. 118, p. 1816-1829, 2009

CASTILHO-MONROY, A. P.; BOWKER, M. A.; MAESTRE, F. T.; RODRIGUEZ-ECHEVERRIA, S.; MARTINEZ, I.; BARRAZA-ZAPEDA, C. E.; ESCOLAR, C. Relationships between biological soil crusts, bacterial diversity and abundance, and ecosystem functioning: insights from a semi - arid Mediterranean environment. **Journal of Vegetation Science**, v. 22, p. 165-174, 2011

CHAER, G. M.; FERNANDES, M. F.; MYROLD, D. D.; BOTTOMLEY, P. J. Shifts in microbial community composition and physiological profiles across a gradient of induced soil degradation. **Soil Science Society American Journal**, v. 73, p. 1327-1334, 2009

DIJKSTERHUIS, G. B.; GOWER, J. C. The interpretation of Generalized Procrustes Analysis and allied methods.**Food Quality and Preference** v. 3, p. 67-87, 1992

DOLÉDEC, S.; CHESSEL, D. Co-inertia analysis: an alternative method for studying species–environment relationships. **Freshwater Biology**, v. 31, p. 277–294, 1994

DRAY, S.; CHESSEL, D.; THIOULOUSE, J. Procrustean Co-inertia analysis for the linking of multivariate datasets. **Ecoscience**, v. 10, p. 110-119, 2003

DUMBRELL, A. J.; NELSON, M.; HELGASON, T.; DYTHAM, C.; FITTER, A. H. Relative roles of niche and neutral processes in structuring a soil microbial community. **ISME Journal**, v. 4, **p.** 337–345, 2009

FITZSIMONS, M. S.; MILLER, R. M.; JASTROW, J. D. Scale-dependent niche axes of arbuscular mycorrhizal fungi. **Oecologia**, v. 158, p. 117-127, 2008

GOWER, J. C. Statistical methods of comparing different multivariate analyses on the same data. In: HODSON, F.R.; KENDALL, D. G.; TAUTU, P.; editors. Mathematics in the archeological and historical sciences. Edinburgh University Press, Edinburgh. pp. 138-149, 1971

GRAYSTON, S. J.; CAMPBELL, C. D.; BARDGETT, R. D.; MAWDSLEY, J. L, CLEGG, C. D.; RITZ, K.; GRIFFITHS, B. S.; RODWELL, J. S.; EDWARDS, S. F.; DAVIES, W. J.; ELSTON, D. J.; MILLARD, P. Assessing shifts in microbial community structure across a range of grasslands of differing management intensity using CLPP, PLFA, and community DNA techniques. **Applied Soil Ecology**, v. 25, p. 63- 84, 2004

HIRST, C. N.; JACKSON, D. A. Reconstructing community relationships: the impact of sampling error, ordination approach, and gradient length. **Diversity & Distribution**, v. 13: p. 361-371, 2007

JACKSON, D. A. PROTEST: A Procrustean randomization test of community environment concordance. **Ecoscience**, v. 2, p. 297-303, 1995

JESUS, E. C.; MARSH, T. L.; TIEDJE, J. M.; MOREIRA, F. M. S. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **ISME Journal**, v. 3, p. 1004-1011, 2009

KANG, S.; MILLS, A. L. Soil bacterial community structure changes following disturbance of the overlying plant community. **Soil Science**, v. 169, p. 55-65, 2004

KUEHNELT-LEDDIHN, E. R. The menace of the herd or Procrustes at large. The Bruce Publishing Company Milwaukee, Auburn. 385 p, 2007

LANDEIRO, V. L.; BINI, L. M.; COSTA, F. R. C.; FRANKLIN, E.; NOGUEIRA, A.; DE SOUZA, J. L. P.; MORAES, J.; MAGNUSSON, W. E. How far can we go in simplifying biomonitoring assessments? An integrated analysis of taxonomy surrogacy, taxonomic sufficiency and numerical resolution in a mega diverse region.**Ecological Indicators**, v. 23, p.366-373.

LEGENDRE, P.; GALLAGHER, E. D. Ecologically meaningful transformations for ordination of species data.**Oecologia**, v. 129, p. 271–280, 2001

LEGENDRE, P.; LEGENDRE, L. *Numerical Ecology*, 3rd English edn. Elsevier Science BV, 516 Amsterdam, 2012

LEGENDRE, P.; BORCARD, D.; PERES-NETO, P. R. Analysing beta diversity: partitioning the spatial variation of community composition data. **Ecological Monographs**, v. 75, p. 435–450, 2005

LISBOA, F. J. G.; CHAER, G. M.; JESUS, E. C.; GONÇALVES, F. S.; SANTOS, F. M.; DE FARIA, S. M.; CASTILHO, A.; BERBARA, R. L. L. The influence of litter quality on the relationship between vegetation and below-ground compartments: a Procrustean approach. **Plant and Soil**, v. 367, p. 551-562, 2012

MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research,** v. 27, p. 209-220, 1967

MERILÄ, P.; LÄMSA- M. M; STARK, S.; SPETZ, P.; VIERIKKO, K.; DEROME, J.; FRITZE, H. Soil organic matter quality as a link between microbial community structure and vegetation composition along a successional gradient in a boreal forest. **Applied Soil Ecology**, v. 46, p. 259 – 267, 2010

MITCHELL, R. J.; HESTER, A. J.; CAMPBELL, C. D.; CHAPMAN, S. J.; CAMERON, C. M.; HEWISON, R. L.; POTTS, J. M . Is vegetation composition or soil chemistry the best predictor of soil microbial community? **Plant and Soil**, v. 333, p. 417-430, 2010

MITCHELL, R. J.; HESTER, A. J.; CAMPBELL, C. D.; CHAPMAN, S. J.; CAMERON, C. M.; HEWISON, R. L.; POTTS, J. M. Explaining the variation in the soil microbial community: do vegetation composition and soil chemistry explain the same or different parts of the microbial variation? **Plant and Soil**, v. 351, p. 355-362, 2012

NEKOLA, J. C.; WHITE, P. S. The distance decay of similarity in biogeography and ecology.**Journal of Biogeography**, v. 26, p. 867–878, 1999

PEIXOTO, R. S.; CHAER, G. M.; FRANCO, N.; REIS JUNIOR, F. B, MENDES, I. C.; ROSADO, A. S. A decade of land use contributes to changes in the chemistry, biochemistry and bacterial community structures of soils in the Cerrado. **Antonie Van Leeuwenhoek**, v. 98, p. 403-413, 2010

PERES-NETO, P. R.; LEGENDRE, P.; DRAY, S.; BORCARD, D. Variation partitioning of species data matrices: estimation and comparison of fractions. **Ecology**, v. 87, p. 2603-2613, 2006

PERES-NETO, P. R.; JACKSON, D. A. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. **Oecologia**,v. 129, p. 169-178, 2001

POMARA, L. Y.; RUOKOLAINEN, K.; TUOMISTO, H.; YOUNG, K. Avian composition co-varies with floristic composition and soil nutrient concentration in Amazonian upland forests. **Biotropica**, v. 44, p. 545-553, 2012

POOS, M. S.; JACKSON, D. A. Addressing the removal rare species in multivariate bioassessments: The impact of methodological choices. **Ecological Indicators**, v. 18, p. 82-90, 2012

POULSEN, A. D.; TUOMISTO, H.; BALSEV, H. Edaphic and florist variation within a 1-ha plot of lowland Amazonian rain forest. **Biotropica**, v. 38, p. 468-478, 2006

POWERS, J. S.; BECKNELL, J. M.; IRVING, J.; PÈRES-AVILES, D. Diversity and structure of regenerating tropical dry forests in Costa Rica: Geographic patterns and environmental drivers. **Forest Ecology & Management**, v. 258, p. 959-970, 2009

SINGH, B. K.; MUNRO, S.; REID, E.; ORD, B.; POTTS, M.; PATERSON, E.; MILLARD, P. Investigating microbial community structure in soils by physiological, biochemical and molecular fingerprinting methods. **European Journal of Soil Science**, v. 57, p. 72-82, 2006

SINGH, B. K.; NUNAN, N.; RIDGWAY, K. P.; MCNICOL, J.; YOUNG, J. P. W.; DANIELL, T. J.; PROSSER J. I.; MILLARD, P. Relationship between assemblages of mycorrhizal fungi and bacteria on grass roots. **Environmental Microbiology**, v. 10, p. 534-542, 2008

SIQUEIRA, T.; BINI, L. M.; ROQUE, F. O.; COTTIENE, K. A metacommunity framework for enhancing the effectiveness of biological monitoring strategies. **PLoS One** v. **7**, e43626, 2012

TEN BERGE, J M. F.; KIERS, H. A. L.; COMMANDEUR, J. J. F. Orthogonal Procrustes rotation for matrices with missing values. **British Journal of Mathematical and Statistical Psychology**, v. 46, p. 119-134, 1993

TER BRAAK, C. J. F.; SCHAFFERS, A. P. Co-correspondence analysis: a new ordination method to relate two community compositions. **Ecology**, v. 85, p. 834-846, 2004

THIOULOUSE, J.; SIMIER, M.; CHESSEL, D. Simultaneous analysis of a sequence of paired ecological tables. **Ecology**, 85, v. 272-283, 2012

TRIVEDI, M. R.; MARECROFT, M. D.; BERRY, P. M.; DOWSON, T. P. Potential effects of climate change on plant communities in three montane nature reserves in Scotland, UK. **Biology & Conservation**, v. 141, p. 1665-1675, 2008

TUOMISTO, H.; POULSEN, A. D.; RUOKOLAINEN, K.; MORAN, R. C.; QUINTANA, C.; CELLI, J.; CAÑAS G. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. **Ecological Applications**, v. 2, p. 352-371, 2003a.

TUOMISTO, H.; RUOKOLAINEN, K.; AGUILAR, M.; SARMIENTO, A. Floristic patterns along 43-km long transect in an Amazonian rain forest. **Journal of Ecology**, v. 91, p. 743-756, 2003

TUOMISTO, H.; RUOKOLAINEN, K.; YLI-HALLA, M. Dispersal, environment, and the floristic variation of western Amazonian forests. **Science**, v. 299, p. 241-244, 2003b

TUOMISTO, H.; RUOKOLAINEN, K. Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. **Ecology**, v. 87, p. 2697–2708, 2006

VAN DER HEIJDEN, M. G. A.; KLIRONOMOS, J.; URSIC, M.; MOUTOGLIS, P.; STREITWOLF-ENGEL, R.; BOLLER, T.; WIEMKEN, A.; SANDERS, I. R. Mycorrhizal fungi determines plant biodiversity, ecosystem variability and productivity. **Nature**, v. 396, p. 69–72, 1998

VINTEN, A. J. A.; ARTZ, R. R. E.; THOMAS, N.; POTTS, J. M.; AVERY, L.; LANGAN, S. J.; WATSON, H.; COOK, Y.; TAYLOR, C.; ABEL, C.; REID, E.; SINGH, B. K. Comparison of microbial community assays for the assessment of stream biofilm ecology. **Journal of Microbiology Methods**, v. 85, p. 190-198.

ZHENG, Y. M.; CAO, P.; FU, B.; HUGHES, J. M.; HE, J. Z. Ecological Drivers of Biogeographic Patterns of Soil Archaeal Community. **PloS One**, v. 8, e63375, 2013

4.CAPÍTULO II:

THE MATCH BETWEEN MICROBIAL COMMUNITY STRUCTURE AND SOIL PROPERTIES IS MODULATED BY LAND USE TYPES AND SAMPLE ORIGIN WITHIN AN INTEGRATED AGROECOSYSTEM

Capítulo publicado como:Lisboa, F. J. G.; Chaer, G. M.; Fernandes, M. F.; Berbara, R. L. L.; Madari, B. E. The match between microbial community structure and soil properties is modulated by land use types and sample origin within an integrated agroecosystem**. Soil Biology and Biochemistry**, v. 78: 97-108, 2014b

4.1. RESUMO

É global a preocupação em adotar medidas para mitigar a degradação de terra causada pelos sistemas de produção agrícola. Uma das estratégias propostas é a substituição de pastagens degradadas por tipo de uso da terra o qual integra três diferentes tipos de uso da terra: agricultura, pecuária, e floresta plantada, coletivamente chamados de iLPF. Contudo, pouco é conhecido sobre as diferenças entre os iLPF e outros tipos de uso da terra em termos de estrutura da comunidade microbiana do solo. Por meio do uso da métrica de associação Procrusteana em ANOVA foi investigado como o tamanho dos resíduos procrusteanos, representando a força de correlação entre variáveis individuais de solo e variáveis individuais de estrutura microbiana, eram particionados em termos de tipo de uso da terra (iLPF; pastagem degradada, pastagem melhorada, fragmento de vegetação nativa), e em termos de posição da amostra de solo dentro do sistema iLPF (área sob maior atuação de árvores, área de maior influência da pastagem, área de transição entre influência de árvores e de influência da pastagem). Foram obtidos indícios de que de fato o tipo de uso da terra pode influenciar mais do que propriedades químicas e variáveis microbiológicas individuais. O sistema ILPF promoveu a dominância de fungos em ambiente de baixo pH e fertilidade. A disponibilidade de P e a variável composta formada por bases trocáveis (Ca^{+2}, Mg^{+2}, K^{+}) foram as propriedades do solo cuja as correlações com variáveis de estrutura microbiana foram mais influenciadas pelo tipo de uso da terra e pela posição da amostra dentro do sistema iLPF. Enquanto a for a de correlação entre variáveis de estrutura microbiana e a disponibilidade de P foi dependente do tipo de uso da terra, a resposta da estrutura microbiana às bases trocáveis foi principalmente afetada pela posição da amostra de solo dentro do iLPF. No geral, os resultados apontam que a heterogeneidade vegetal introduzida por meio do plantio de árvores em meio a pastagem dentro dos sistemas iLPF é um importante regulador das respostas da comunidade microbiana às mudança ambientais, e pode ser uma das formas de aumentar a sustentabilidade em agroecosistemas tropicais.

Palavras-chave: Manejo. Agroecosistemas. Relacionamento. Métrica de associação Procrusteana.

4.2. ABSTRACT

It is of global concern to adopt measures to mitigate land degradation caused by agricultural production systems. One of the strategies proposed is to replace degraded pastures with agrosilvopastoral systems which integrate three different land-use types: crop production, livestock pasture and forestry plantation (denoted iCLF). However, little is known about the differences between iCLF and other land use types interms of soil microbial community structure. Distance matrices based on individual soil chemical properties and individual soil microbial variables were correlated by Procrustes analysis and these relationships yielded vectors of residuals depicting these correlations (matches). These vectors were used as univariate response variables in an ANOVA framework in order to investigate how the match sizes (the strength of correlation/covariance) between individual soil chemical properties and individual soil microbial variables vary across land use types (levels: iCLF; degradated pasture; improved pasture; and anative cerrado fragment) and also across sample origin within iCLF (levels: soil samples under more influence of the exotic tree forest stand; soil samples under influence of the pasture; samples within the transition between the forest stand and the pasture). We were able to obtain insights into the fact that the land use distinction can be driven by more than just individual soil chemical and microbial variables. The integration of crop, livestock and forestry promoted a dominance of fungi in this low fertility and low pH environment. P availability and the composite variable exchangeable base cations (Ca^{+2}, Mg^{+2}, K^{+}) were the soil properties whose strengths of correlation (match sizes) with individual microbial variables were the most affected by land use type and sampling origin within iCLF. While the strength of the correlation between soil microbial structure variables and P availability was typically land use type dependent, the response of the microbial structure to exchangeable base cations was mainly affected by the sample origin within iCLF. Finally we concluded that increases in the heterogeneity of vegetation within integrated crop, pasture and forestry systems are an important driver of microbial community response to environmental changes, and may be one means by which to increase the sustainability of tropical agroecosystems.

Keywords: Management. Tree-basedsystems. Agroecosystems. Relationships. Procrustes association metric

4.3. INTRODUCTION

Global concern with farmland degradation, usually associated with soil carbon loss, has led numerous countries to seek management strategies aimed at the restoration and sustainable use of such areas. In Brazil, special attention is being paid to integrated crop-livestock-forest systems (iCLFs) for replacing pastures in different stages of degradation. Approximately 12% of the Earth's land surface is covered by agricultural crops, 33% is intended for livestock, and 15% supports exotic forest species (Giraldo et al., 2011). Pastures accumulate large quantities of carbon in the topsoil layers due to the profusion of fine roots from grasses but produce relatively less recalcitrant substrates compared to forest systems. This favors organic matter mineralization by stimulating a microbial structure with a higher activity (Bardgett and McAlister, 1999), consequently inducing higher soil carbon losses (Millard and Singh, 2009). A recent study showed that vegetation homogenization generated by converting natural forests into pastures may be accompanied by homogenization of the microbial communities (Rodrigues et al., 2013), most likely due to a reduced diversity of good quality substrates per soil volume (Lamb et al., 2010). In contrast, the introduction of tree species may promote microbial diversity when converting pastures into exotic species forests (Carson et al., 2010).

It is believed that changes in microbial community structure generated by modified land management and land use type can be related to the soil switching from carbon source to carbon sink or vice versa. Still, it is suggested that land use types considered to be more conservative regarding organic matter mineralization tend to exhibit a microbial structure with lower activity (Bardgett and McAlister, 1999). Within this formulation, iCLF systems, which combine crop production, managed pasture and forest species, are designed to exhibit a microbial structure distinct from that of degraded pastures via plant physiological heterogenization of the landscape. However, soil microbial community structure is rarely investigated in essentially agrosilvopastoral systems such as iCLF systems, especially in the tropics (Lacombe et al., 2009; Vallejo et al., 2012). Thus, we do not have a large body of evidence that iCLF systems may be more carbon conservative.

Changes that occur in vegetation composition and management type due to land use type conversion are responsible for most of the variation that occurs in chemical and physical soil properties. In turn, these changes tend to correlate with variation in the microbial community, linking the changes above and below the soil surface (Lisboa et al., 2012; Mitchell et al., 2010). However, the extent to which this link between soil chemical variables and the phenotypic structure of the microbial community is partitioned among different land use types, as well as how the man-generated plant heterogeneity introduced by the forest component in the integrated crop-livestock-forest (iCLF) is able to differentiate it from others land use types, remain unaddressed questions.

In this study, we started with the hypothesis that introducing iCLF as replacements for degraded pastures leads to a change in the response of phenotypic composition of the soil microbial community to individual soil chemical variables. First we accessed the individual responses of soil and microbial phenotypic variables to land use type in three different scenarios: 1) considering all samples in the iCLF, 2) considering only samples from the centre of the pasture component of the iCLF; 3) considering only samples from the forest stand within the iCLF. For the main point in this study, i. e. how the matches/effects of individual soil chemical variables on the individual microbial variables are partitioned by land use type and sampling origin within iCLF we used features from an uncommon statistical approach in soil microbial ecology studies called Procrustes analysis (Gower, 1971). Similar to the

somewhat traditional Mantel test, Procrustes analysis is a correlative multivariate approach. However the correlation in Procrustes analysis is reached through rotation and translation, seeking for the "best" fit that depicts the minimal residual difference between homologues coordinates of two or more matrices under analysis (Peres-Neto and Jackson, 2001, Lisboa et al.,2014). These homologues coordinates are nothing but the rows (sites, samples) of the matrices under analysis so that low residuals stand for strong matches/effects whereas high residual differences mean weak matches/effects. Procrustes has the feature of providing the matches among all homologues coordinates of matrices under analysis in a vectored form sometimes called Procrustean associated metric (Lisboa et al., 2014). Thus this vector can be retained for using in downstream statistical approaches in order to investigate the consistencies in the size of the matches across different environmental predictors. In the present study we investigated the consistency in size matches from Procrustean association metric between distance matrices based on individual soil chemical variables and soil microbial variables in an ANOVA framework having the land use type as factor (first ANOVA) and sample origin within iCLF as factor (second ANOVA).

4.4. MATERIALS AND METHODS

4.4.1. Study area

The study was conducted in one of the 203 technology reference units in iCFL (http://www.cnpgl.embrapa.br/nova/silpf) of the Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária - Embrapa) on the Boa Vereda farm located in the municipality of Cachoeira Dourada, Goiás State, Brazil. The study site is located at 18°27'43.19"S, 49°35'58.53"W at an altitude of 484 m above sea level, on a clay (603 g kg^{-1}) Rhodic Ferralsol (Latossolo Vermelho Acriférrico típico (Brazilian Soil Classification System) or Anionic Acrustox (Soil Taxonomy)) with slopes between 0 and 15% and a mean annual rainfall of 1,350 mm (Brasil, 1983). Four land-use types were assessed in this study (Fig. 1):**1)** an iCLF system; **2)** improved pasture with remnants of dry forest natural vegetation (native trees); **3)** degraded pasture; and **4)** a native cerrado fragment (savannah-like) exhibiting 'cerrado denso' (dense tree savannah) vegetation.

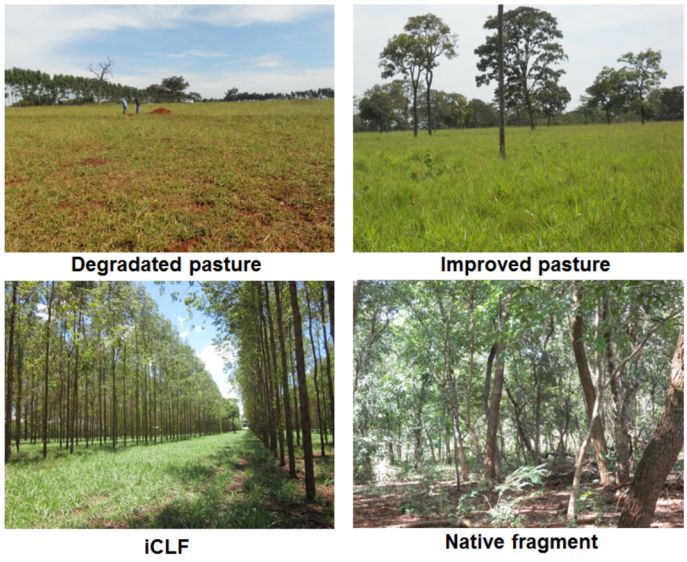


Figure 1. Land use types investigated in this study. iCLF: integrated crop-livestock-forest system; Native fragment: native cerrado (savannah – like) fragment.

4.4.2. History of land use types

Originally, all of the sites studied were covered with 'cerrado' vegetation, within which they represented forest formations of dry forest, 'cerradão' (woodland), and 'cerrado denso'. All of the areas, except original forest, had been deforested for more than 30 years and were maintained as pasture until recently.

iCLF: In 2009, the iCLF system was implemented with three rows of eucalyptus trees per stand, using 476 trees per hectare (ha), in a total area of 14.7 ha with the following management sequence: in August, the soil was plowed with a disc harrow at a cutting depth of 25 cm, and lime was incorporated into the soil. Between October and November, the soil was prepared for planting soybean (*Glycine max* L. variety BRSGO 8360) and eucalyptus (*Eucalyptus urograndis*) with a leveling disc harrow. In October 2010, the soil was prepared again for planting corn (*Zea mays* L.) intercropped with brachiaria grass (*Urochloa brizantha*). After harvesting the corn, the soil was not mechanically turned anymore, and the brachiaria grass developed into pasture between the eucalyptus rows. The soybean crop received 300 kg ha^{-1} 04-30-10 (NPK) + Zn formula fertilizer, the corn received 300 kg ha^{-1}08-

30-10 + Zn and the eucalyptus crop received 150 and 10 g plant^{-1} of 08-30-10 + Zn and boric acid, respectively. As maintenance fertilizer, the eucalyptus received 200 kg of simple superphosphate broadcasted and 15 g boric acid, and the pasture between the eucalyptus rows received 100 kg urea ha^{-1} and 100 kg ha^{-1} monoammonium phosphate annually.

Improved pasture: Before the pasture was restored, it was in a situation of abandonment. In 2008, the site was restored, starting with a disc harrow. Then, lime was applied and incorporated into the soil with a leveling harrow, followed by planting *Brachiaria* grass that continues to grow on the site.

Degraded pasture: This land use type occupies an area of 10.6 ha and comprises kikuyu grass (*Pennisetum clandestinum*) and Surinam grass (*Urochloa decumbens*). Lime and fertilizer were never applied.

4.4.3. Sampling design

The samples were collected in March 2012. Five sample modules were established within each land use type 50 m apart from each other diagonally. Except for native cerrado fragment (NF), where eight samples were collected without using sample modules, three sampling points comprising six cores taken randomly (5 x 5 cm) were collected from each sample module within DP, IP and iCLF. Thus, the total number of samples for each land use type was 15 (DP, IP and iCLF) and 8 (NF), which gave a sample size of 53. Specifically for the man-generated plant heterogeneity created after the introduction of exotic tree species stands (*E. urograndis*), within the pasture (main feature of iCLF), each sample within a given sample module was coming from a different origin. Here, a sample origin designation similar to that employed by Vallejo et al. (2012) was used; the sampling points taken within the tree component of the iCLF system were named **canopy**, the sampling points taken from the crown projection were named **transition**, and finally, the sampling points taken from the center of the pasture component of the iCLF system were named **outside.** So, the total 15 samples within iCLF could be divided according to the origin: canopy (5), transition (5) and outside (5).

4.4.4. Soil variables analysis

Soil pH was determined in water and potassium chloride (KCl) using potentiometry (Thomas, 1996). Phosphorus (P, mg dm^{-3}), calcium (Ca^{2+}, $mmol_c$ dm^{-3}), magnesium (Mg^{2+}, $mmol_c$ dm^{-3}) and potassium (K^+, mg dm^{-3}) were extracted by a dilute solution of strong acids (0.05 mol L^{-1} HCl + 0.0125 mol L^{-1} H_2SO_4; Mehlich I) as described by Kuo (1996). Phosphorus was determined by the colorimetric method (Embrapa, 2009), Ca^{2+} and Mg^{2+} by atomic spectroscopy and K^+ by flame emission spectrometry (Wright and Stuczynski, 1996). Soil organic matter (SOM, mg g^{-1}) was determined by the Walkley-Black method (Nelson and Sommers, 1996) without external heating, using sulfuric acid (H_2SO_4) to create internal heat for the reaction. Soil moisture was determined gravimetrically (mg g^{-1}) and soil bulk density (g cm^{-3}) using the short (5 cm) core method (Grossman and Reinsch, 2002).

4.4.5. Microbial analysis

Phospholipid fatty acid (PLFA) analysis was used to assess the changes in the phenotypic structure of the microbial community because this method has been shown to be effective in discriminating changes in land-use type (Cao et al., 2010; Diedhiou et al., 2009; Kasel et al., 2008). Despite its low resolution, this approach has the advantage of allowing for quantification of different and important microbial groups and indices related to soil function, such as fungi and bacteria (Frostegård et al., 2011; Wixon and Balser, 2013). To obtain the

lipid profile of the soil microbial communities, we followed the method described by Fernandes et al. (2011).

The fatty acid methyl esters (FAMEs) were separated by a gas chromatograph with flame-ionization detector (Clarum 500, PerkinElmer) using a capillary column (5% biphenyl-95% dimethylpolysiloxane, 25 to 30 m) with the following program: 5ºC/minute, from 120 to 270ºC. The injector and detector temperatures were 250ºC and 280ºC, respectively. The chromatogram peaks for each sample were identified by comparing the retention times generated by commercial standards (FAME 37 47885-U and BAME 24 47080-U Sigma-Aldrich).

The area of each peak within the sample was calculated relative to the total area of the chromatogram for obtaining the percentage of FAME within each sample (Fernandes et al., 2011). Twenty informative peaks were common for all samples and then used in the analyses, the results being expressed in percentage mol (row matrix normalization of raw values). The profile of all of the FAMEs (20 in total) was used as a surrogate phenotypic structure of the microbial community. Within the total profile different markers were pooled in order to obtain proxies for the following microbial groups. Gram (+) bacteria (i15:0, a15:0, 15:0, i16:0, a16:0, i17:0, a17:0, 17:0) ; Gram (-) bacteria (16:1ω7c, 18:1ω7c, 18:1ω9c, cy17:0, and cy19:0) ; bacteria (Gram (+) plus Gram (-)) fungi (18:2ω6,9); arbuscular mycorrhizal fungi-AMF (16:1ω5c); and actinomycetes (10Me17:0, 10Me16:0). The fungi: bacteria ratio (F:B ratio) was obtained by dividing the percentage mol of the 18:2ω6,9 FAME by the sum of the percentage mol of the bacterial FAMEs. The ratios between FAMEs cy17:0 and cy19:0 and their respective FAME precursors, 16:1ω7c and 18:1ω7c, were used to measure microbial stress (Frostegård et al., 1993; Olsson, 1999; Zelles, 1999).

4.4.6. Data analysis

All of the analyses were conducted using packages available for the R statistical program (*R* Core Development Team, 2013). For an initial analysis of the effect of the different land-use types on the overall phenotypic variation of the microbial community, the FAMEs profile (percentage mol) was subjected to PERMANOVA (Anderson, 2001). To evaluate whether the differences in microbial structure remained consistent across samples from different sampling points of the iCLF system, we conducted three distinct PERMANOVA´s: 1) considering all of the samples, independent of their origin within the iCLF; 2) only considering the samples from the center of the pasture within the iCLF (outside); and 3) only considering the samples obtained in the tree stand within the iCLF (canopy). PERMANOVA was conducted using the *adonis()* function of the vegan package (Oksanen et al., 2013).

We used between-group analysis (BGA) to assess visually and statistically the differences (or similarities) in microbial structure across land-use types and the potentially predominant microbial groups. Like a discriminant analysis, BGA uses scores from ordination methods as response variables and a factor with different levels as a categorical variable searching for the best combination of variables that allows the scores of the ordination axes to maximize the relationship for within- and between-level variation. In this study the scores were coming from a PCA based on a matrix of microbial groups and indices whereas the factor was the land use type (DP, IP, NF and iCLF) (Thioulouse et al., 2012). Similar to PERMANOVA, three distinct comparisons among land types using BGA were conducted based on the sample origin within iCLF: 1) considering the samples within the iCLF system; 2) only samples from exotic tree stands within iCLF (canopy); 3) only samples from the pasture within iCLF (outside)**.** The BGA analyses were conducted using the *bca()* function and the overall statistical difference among land use types was accessed by permutations

using randtest.between*()*. These functions are available in the ade4 package (Chessel et al., 2004). Final BGA plots were customized using Microsoft PowerPoint.

To test whether the correlations between the individual soil variables and the microbial groups and indices differ as a function of land use type, we utilized the following procedure. First, individual soil variables — except for exchangeable base cations (Ca^{2+}, Mg^{2+} and K^{+}), as the sum of those was considered (EBC) — and each of the different microbial groups and indices based on percentage mol FAMEs were log(x + 1)-transformed and used to construct dissimilarity matrices (Euclidean). In the second step, each microbial distance matrix was related to each matrix of the individual soil variables to answer to the general question of whether the environmental distances based on individual soil variables are significantly related to the environmental distances based on different microbial groups and indices. Two analogous approaches were used: the partial Mantel test (Legendre and Legendre, 2012) and the partial PROTEST (Peres-Neto and Jackson, 2001), where both isolated the effect of the all other soil covariates before conducting the relationship analysis. Finally, since PROTEST (Procrustes analysis) has the feature of providing the relationship between matrices in a vectored form (Peres-Neto and Jackson, 2001) we used the vectors from significant results of the partial PROTEST in two different a one-way ANOVAs. Thus, the vectors representing significant relationships between the distance matrices based on the individual variables and the distance matrices of the microbial groups and indices were used as the response in two different ANOVAs, the first one by using land use type as factor (DP, IP, NF and iCLF). Finally, the vectors that were significantly affected in the ANOVAs were subjected to a test of means (least significant differences - LSD, 95% confidence, with Bonferroni correction) to assess how the magnitude of the effects (matches) of the soil variables on microbial structure varies between the land-use types.

The same approach described in the paragraph above was followed within the iCLF system, i.e., the significant relationships were subjected to ANOVA, but now with sample position in the iCFL (canopy, transition, outside) as a factor. Thus, one can assess if the heterogeneity generated by introducing exotic tree species into the pasture is able to affect the extent to which different soil variables influence microbial structure. The partial Mantel test and partial PROTEST were conducted using the vegan package (Oksanen, 2013), whereas the test of means was conducted using the agricolae package in R (Mendiburu, 2014).

4.5. RESULTS

4.5.1. Individual soil variable correlations

The correlation plots between individual soil variables showed that the use of samples from different origins within the iCLF may be influential. By using as rule of thumb the value 0.50, the general trend was that in the scenarios using all and only samples from the planted tree stands the correlations were positive and relatively strong, particularly between soil variables K, Ca and Mg and between these and pH (Fig. 2ab). On other hand, the use of samples coming from the centre of the pasture within iCLF gave more negatively and moderately strong relationships, particularly the correlation of moisture and bulk density with the exchangeable base cations such as K, Ca and Mg (Fig 2abc).The pH was positively correlated with all exchangeable base cations across all sampling origin scenarios (Fig. 2abc). Moreover it is relevant to highlight that within the exchangable base cations the correlations were positive from moderately to strong across the three sampling origin scenarios investigated (Fig. 2abc), which justifies the use of the composite variable EBC. The soil organic matter related weakly with other soil variables across all the different scenarios (Fig 2abc).

4.5.2. Land use ordination based on soil variables

The principal component ordination based on soil variables was done to give an intuitive picture of general differences among land use type across the different scenarios of sampling origin within the iCLF. Across all scenarios the overall trend was the grouping of iCLF and degradated pasture (DP) and their distance from the improved pasture (IP) and native fragment (NF) (Fig. 3abc). This general pattern seemed to be mainly driven by the contrasts in terms of the variable bulk density, which was higher in iCLF and DP than in the IP and NF (Fig. 3abc). In the first two scenarios of sampling origin within iCLF (all samples and canopy) the land use types iCLF, DP and IP showed to have low soil fertility when compared to NF (Fig. 3ab). However when only samples from the centre of the pasture within the iCLF (outside) were considered the IP had higher fertility than iCLF and DP (Fig. 3c). It is important to notice that only soil variables related significantly (P<0.05) with ordination axes are showed. Thus, SOM did not have a great contribution for the land use type discrimination as it was not related to the axes across all scenarios investigated.

4.5.3. Individual variables - based pairwise dissimilarity as affected by land use and position sampling within iCLF (PERMANOVA)

With the exception of the Gram (+) bacterial profile, the PERMANOVA indicated significant differences between the land-use types for all of the microbial groups and indices when considering all of the samples of the iCLF system (Table 1). Gram (+) bacteria, Gram (-) bacteria, actinomycetes, and the cy19/18:1ω7 ratio were unable to differentiate the land-use types when the samples from the forest component (canopy) were considered (Table 1). However, when the pasture replicates were considered (outside), the land-use types differed based on the Gram (-) bacterial profile (F=3.629;P< 0.05). Additionally, the land-use types were significantly different regarding the 16:1ω5c marker for AMF only when PERMANOVA considered all of the samples (F=6.701; P < 0.01) and those from the forest component of the iCLF (F=4.157; P < 0.05). Interestingly, the land-use types remained distinct for the general microbial, bacterial, and fungal profiles and for the F:B and cy17:0/16:1ω7c ratios, regardless of microbial sample position (Table 1).

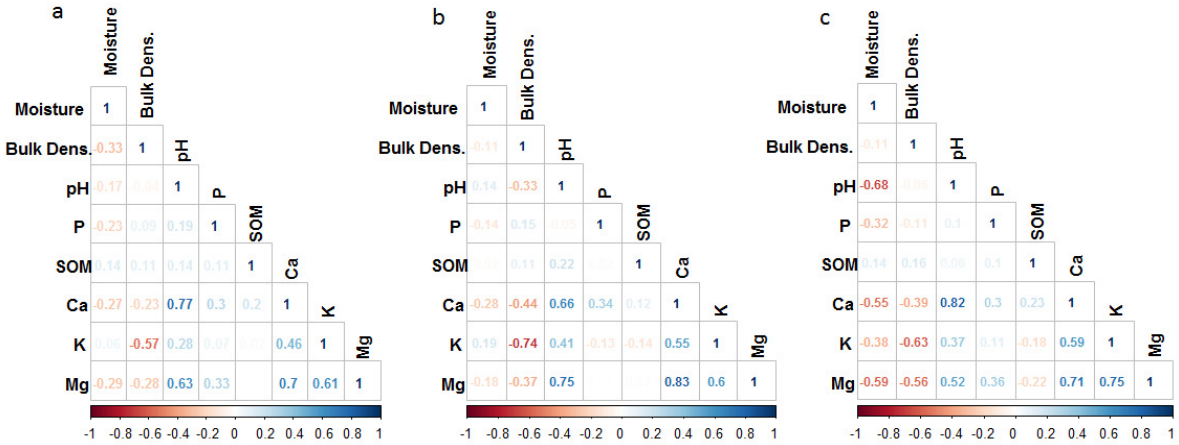


Figure 2. General view of the Soil variables correlation in three scenarios of sample origin within the integrated crop-livestock-forest (iCLF): a) all samples within iCLF; b) only canopy samples; c) only outside samples). Weakness of number colors stands for relationship magnitude while the color type informs on the nature of the relationship: more blue (positive), more red (negative). Since the primary objective is to visualize the magnitude of relationships and patterns the significance is not shown. SOM (soil organic matter); Bulk dens. (bulk density). All variables were log transformed before correlation analysis.

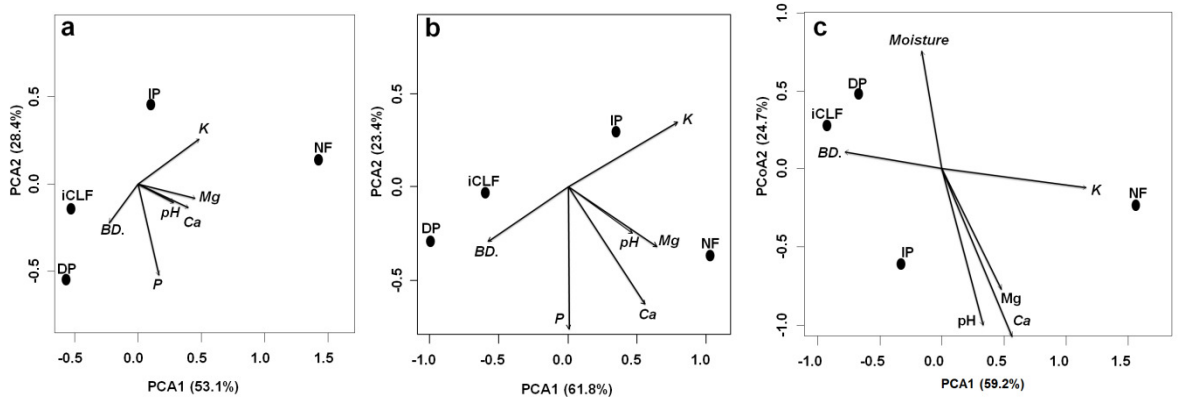


Figure3. PCA based on soil variables. Three scenarios were used: a) PCA by using all Samples within iCLF; b) PCA by using five samples coming from planting tree line within iCLF (canopy); c) PCA by using samples from the centre of the pasture within the iCLF (outside). Here the centroids of each land use type (iCLF: integration crop-livestock-forest; DP: degradated pasture; IP: improved pasture; NF: native cerrado fragment) have been shown in order to get a more clear representation in a reduced space. Arrows are soil variables that are significantly related to PCA axes (P<0.05). BD (bulk density).

Regarding the soil variables, pH (H_2O and KCl), available phosphorus (P), and soil organic matter, these variables were unable to discriminate the land-use types across all scenarios of sampling within iCLF (Table 1). In contrast, moisture, bulk density, and exchangeable based cations were able to differentiate the land-use types regardless of the samples' position within iCLF system (Table 1).

4.5.4. Between group analyses based on microbial variables across different sampling scenarios

The PCA generated to run the BGA was based on a matrix of microbial groups and indices. The BGA permutations reinforced the PERMANOVA results for soil microbial variables by showing that overall differences among land use types were not random. This was valid for all of the samples (observed value: 0.44;*P*< 0.001); the canopy samples (observed value: 0.53; *P*< 0.001); and the samples outside of the canopy (observed value: 0.49;*P*< 0.001).

Overall the BGA shows that across all scenarios of sampling origin within iCLF there was a trend of grouping iCLF and IP contrasting with other land use types (DP and NF) (Fig. 4a-c). Moreover, the distinction of iCLF and IP from DP and NF seems to be driven by contrasts between Fungal and Bacterial variables along the highest variation axis (Fig. 4abc).

4.5.5. The access and significance of the Partial Protest matches

Despite partial PROTEST to be the ground of the present study we have ran partial Mantel just like a "devil's lawyer" in order to check for consistency of the effect of individual soil variables on microbial variables. Thus, we did not intent to make formal comparisons between these two approaches since it has been done elsewhere. Considering the results of the partial PROTEST, overall microbial structure profile (PLFA), overall bacterial profile, Gram (-) bacteria, fungi, F:B and cy19:0/18:1ω7c ratios were affected by all of the soil variables included in the study (Table 2). It is worthwhile to notice that the exchangeable base cations composite variable (EBC) was the unique to show consistent significant effect on microbial community and indices throughout both partial Mantel and partial Procrustes (Table 2). In general, the matches of the 18:2ω6,9 FAME (fungi) and the F:B ratio to soil variables were similar.

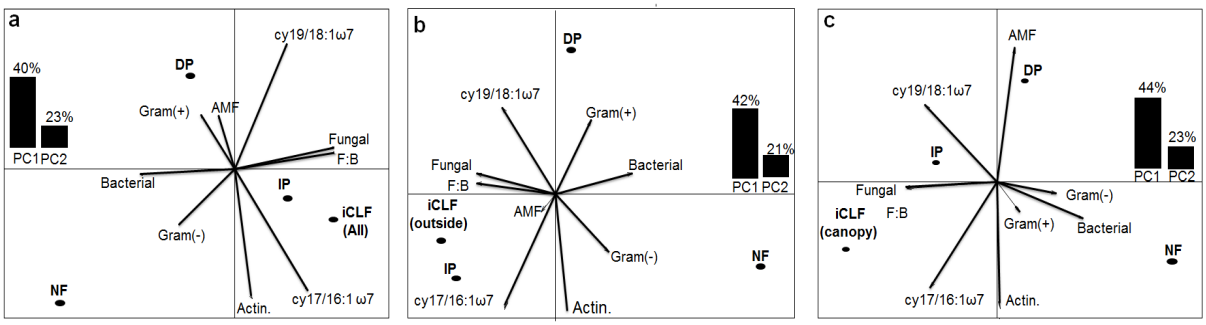


Figure 4. Between Group Analysis(BGA, Thioulouse et al., 2012) using scores from a PCA based on a matrix of microbial groups and indices as response. Land use type was the explanatory categorical variable where each level is represented by the centroid. Arrows stand for microbial variables as in a biplot (no P cutoff used). Figure a) shows the BGA results considering all samples in the integrated crop-livestock-forest land use (iCLF); b) accounts for BGA using only samples from the center of the pasture within iCLF (outside); c) illustrates the BGA analysis conducted with only samples from the base of tree stands within iCLF (canopy). The full circles represent the centroid of the land use types: iCLF; IP (improved pasture); DP (degraded pasture); NF (native cerrado fragment); Actin. (Actinomycetes); AMF (arbuscular mycorrhizal fungi); Bacterial (Gram (+) + Gram(-)); F:B (Fungal:Bacterial ratio).Filled bars indicate the variation explained by the first two principal axes.

Table 1.PERMANOVA.Analyses using microbial groups and indices (%mol FAMEs) and individual soil variables as responses and land-use type as a categorical predictor (4 levels$^{\Delta}$).

	Samples (All)[1]		Samples (Canopy)[2]		Samples (Outside)[3]	
	F	*P*	*F*	*P*	*F*	*P*
PLFA variables						
PLFA profile	14.57	***	6.368	***	8.387	***
Gram(+)	2.847	ns	0.730	ns	2.701	ns
Gram (-)	5.576	***	2.448	ns	3.629	*
Bacterial	30.58	***	15.23	***	13.65	***
Fungal	32.12	***	15.87	***	12.66	***
Actinomycetes	8.465	**	2.839	ns	3.177	ns
AMF	6.709	**	4.157	*	1.173	ns
F:B ratio	32.82	***	16.47	***	12.01	***
cy17/16:1ω7	18.70	***	8.424	**	8.781	**
cy19/18:1ω7	11.68	***	0.804	ns	1.607	ns
Soil variables						
Moisture	19.31	***	12.13	***	4.952	*
Bulk density	22.36	***	10.766	***	7.978	**
pH (H_2O)	1.921	ns	0.863	ns	1.234	ns
pH (KCl)	2.868	*	0.571	ns	0.523	ns
Exchangeable base cations	18.43	***	11.62	***	17.21	***
P available	2.600	ns	0.555	ns	1.473	ns
Soil organic matter	1.284	ns	1.651	ns	0.819	ns

Δ T1: integration crop-livestock-forest (iCLF); T3: degraded pasture; T4: improved pasture, T5: native cerrado fragment.1Analyses considering all of the samples from the integrated system (iCLF) regardless of position. 2Analyses conducted considering only the samples from the forest planting within the iCLF system (Canopy). 3 Analyses conducted considering only the samples from the center of the pasture within the iCLF system. All of the PERMANOVAs were conducted using the Euclidean distance of the log(x +1)-transformed data. *P < 0.05; **P< 0.01; ***P< 0.001. ns: not significant (P> 0.05).: distance matrix based on the sum of Ca^{2+}, Mg^{2+} and K^{+}.

4.5.6. The analysis of the partial PROTEST matches in ANOVA framework

Despite the 38 significant relationships of soil variables with microbial groups and indices identified by the partial PROTEST (Table 2), only 13 were detected by ANOVA as being significantly affected by variation in land-use type (Table 2). None of the significant effects of the moisture and of the bulk density on microbial variables given by partial PROTEST were affected by land use type (Table 2). Within the set of the significant effect of pH on microbial variables, the effects on Bacterial, F:B ratio, and total PLFA profile had their magnitude partitioned by land use type (Table 2).

The post hoc test revealed these land use affected matches were generally lower in the iCLF system and degraded pasture compared to the effects in the improved pasture (IP) and in the native cerrado fragment(NF) (Fig. 5).For the exchangeable base cations variable matches affected by land use type (Table 2), the effect of this composite variable on the Gram (-) only differed between DP and NF land uses (Fig. 5).On other hand the effect of EBC on cy19/pre were higher in the DP and iCLF when compared to NF and IP land use types (Fig. 5).

Table 2. The significant relationships (partial Mantel and partial Protest) between Euclidean distance matrices based on individual soil variables and Euclidian distance matrices based on microbial groups and indices. The significant Procrustean relationships (partial Protest) in a form of vectors were used as response in a one-way ANOVA framework (factor: land use type with four levels$^{\Delta}$).

Relationships(matches)	Partial Mantel^{Euc}	Partial Protest^{Euc}	Land use (ANOVA)	
			F	*P*
Moisture x total PLFA	0.25***	0.44**	1.030	0.395
Moisture x Gram(-)	0.16**	0.30*	0.576	0.633
Moisture x Bacterial	0.17**	0.45***	0.703	0.555
Moisture x Fungal	ns	0.47***	0.974	0.413
Moisture x AMF	ns	0.32*	2.296	0.085
Moisture x F:B ratio	ns	0.47***	0.749	0.528
Bulk dens. x PLFA	0.19**	0.46***	0.443	0.749
Bulk dens. x Gram(-)	0.15**	0.35**	0.381	0.767
Bulk dens. x Fungal	ns	0.22*	0.892	0.463
Bulk dens. x F:B ratio	ns	0.28*	0.923	0.437
Bulk dens x cy19/pre	0.25**	0.48***	0.118	0.949
pH(H_2O) x total PLFA	ns	0.38**	2.510	0.070
pH(H_2O) x Gram(-)	ns	0.28*	2.039	0.095
pH(H_2O) x Bacterial	ns	0.32*	3.841	0.015*
pH(H_2O) x F:B ratio	ns	0.29*	3.576	0.020*
pH(KCl) x total PLFA	ns	0.37**	2.841	0.041*
pH(KCl) x Gram(-)	ns	0.29*	2.768	0.051
pH(KCl) x Bacterial	ns	0.32*	3.818	0.015*
pH(KCl) x cy19/pre	ns	0.30**	2.051	0.119
EBC x total PLFA	0.34**	0.57***	1.004	0.399
EBC x Gram(-)	0.20**	0.39**	3.224	0.030*
EBC x Bacterial	0.20***	0.44***	0.754	0.525
EBC x Fungal	0.20**	0.40***	1.852	0.150
EBC x F:B ratio	0.20**	0.40**	1.573	0.208
EBC x cy19/pre	0.10*	0.35**	3.564	0.020*
P x total PLFA	ns	0.031*	3.693	0.017*
P x Bacterial	ns	0.28*	3.913	0.013*
P x Fungal	0.14*	0.34**	3.251	0.029*
P x Actin.	ns	0.33*	3.825	0.015*
P x AMF	ns	0.31*	6.227	0.001**
P x F:B ratio	0.14*	0.34**	3.225	0.030*
SOM x total PLFA	0.18**	0.46**	1.657	0.188
SOM x Gram(-)	ns	0.33**	1.823	0.154
SOM x Bacterial	ns	0.37**	1.177	0.328
SOM x Fungal	ns	0.42***	0.353	0.787
SOM x Actin.	ns	0.31*	4.136	0.010*
SOM x F:B ratio	ns	0.42***	0.334	0.800
SOM x cy19/pre	ns	0.31*	0.640	0.593

^{Euc} indicates "Euclidean", which was the resemblance measure used for building the distance matrices. Prior to calculation of dissimilarities, the variables related were log(x+1)-transformed. PLFA: Fatty acid profile. Likewise, the match/effect of each soil variable distance matrix was tested after accounting for covariance with other soil variables. $^{\Delta}$ Land use factor levels are: DP (degradated pasture); IP (improved pasture); NF (native cerrado fragment); iCLF (integration crop-livestock-forest). "EBC" is a composite variable referring to exchangeable base cations (Ca^{2+}, Mg^{2+} and K^{+}). Actin (actinomycetes); AMF (arbuscular mycorrhizal fungi); Bacterial (Gram(+) + Gram(-)); F:B (Fungal:Bacterial ratio); cy19/pre(cy19/18:1ω7); SOM(soil organic matter). Here partial Mantel has been run to verify match consistencies given by partial PROTEST (main analysis). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

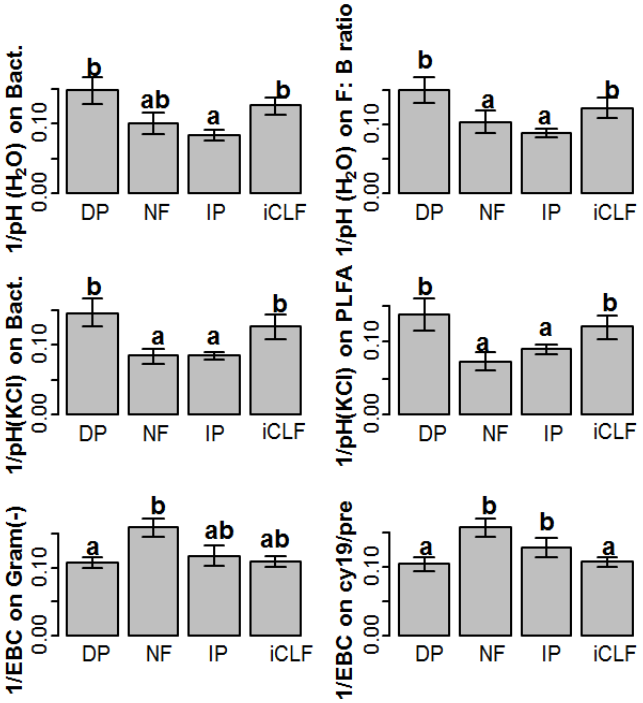


Figure 5. Mean test (LSD at 5% of significance; *P* value adjusted by Bonferroni) showing the variation in the match sizes of the pH and exchangeable base cations (EBC)to soil variables on microbial groups and indices as affected by the land use types. All these matches are coming from partial PROTEST (Procrustes analyses) based on Euclidian of log(x+1) dissimilarities matrices of individual variables. Here, the only effects of EBC and pH shown are those that were considered to be affected significantly by land use type in the ANOVA (Table 2). IP (improved pasture); DP (degraded pasture); NF (native cerrado fragment), integrated crop-livestock-forest (iCLF). PLFA (general microbial profile-20 peaks); Bacterial (Gram (-) + (Gram (+)); F:B (Fungal: Bacterial ratio). The expression 1/match is used because these effects are Procrustes association metrics representing the matching between individual soil chemical and soil microbial variables, which are inversely proportional to the level of relationship (or effect), so a higher value indicates a lower effect. Standard errors (SE) are shown in each bar.

The significant matches between P and microbial variables were all affected by the land use type (Table 2). A clear pattern was found given that the effect of the P on microbial structure variables was consistently low in iCLF (Fig. 6).

4.5.7. The partitioning of the partial Procrustes matches within iCLF (sampling origin as ANOVA factor)

The 38 significant relationships provided by the partial PROTEST from Table 2 were again subjected to ANOVA, but now considering the sample position within the iCLF as a three level factor (canopy, transition, outside).Only the effects of the exchangeable base cations variable (EBC) on microbial variables were affected by the sample origin, namelyits effects on the Gram (-), bacteria, fungi, F:B ratio, and cy19:0/18:1ω7c ratio (Table 3). Post hoc test revealed that all of these effects were quantitatively higher in the (canopy) than in the (transition) and (outside) (Fig. 7).

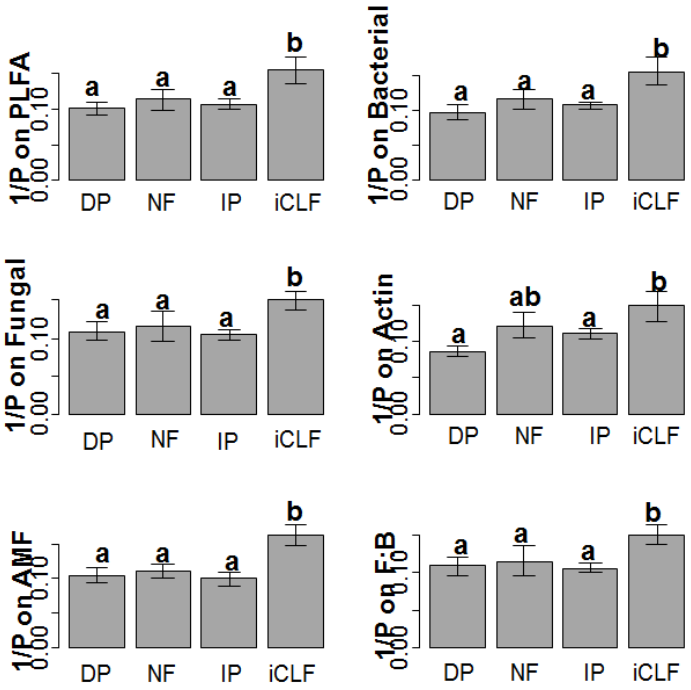


Figure 6. Mean test (LSD at 5% of significance; *P* value adjusted by Bonferroni) showing the variation in the match sizes of the Pto the soil variables on microbial groups and indices as affected by the land use types. All these matches are coming from partial PROTEST (Procrustes analyses) based on Euclidian of log(x+1) dissimilarity matrices of individual variables. Here, the only effects of Pon microbial variables shown are those that were considered to be affected significantly by land use type in the ANOVA (Table 2). IP (improved pasture); DP (degraded pasture); NF (native cerrado fragment), integrated crop-livestock-forest (iCLF). PLFA (general microbial profile-20 peaks); Bacterial (Gram(-) + (Gram(+)); F:B (Fungal:Bacterial ratio); Actin (actinomycetes); AMF (arbuscular mycorrhizal fungi). The expression 1/match is used because these effects are Procrustes association metrics representing the matching between individual soil chemical and soil microbial variables, which are inversely proportional to the level of relationship (or effect), so a higher value indicates a "lower effect". Standard errors (SE) are shown in each bar.

4.6. DISCUSSION

The word "heterogeneity" used here is not related to the gradient of plant diversity across all land use types investigated. It stands for the deliberate heterogeneity (i.e. a man-driven landscape display) which characterizes the integrated crop-livestock-forest land use type (iCLF).This man-driven heterogeneity is mainly caused by the introduction of regular tree stands in a pasture followed by soil management. Thus this study has tried to raise relevant insights on how this introduced heterogeneity within iCLF makes this land use type distinct from others, not only by investigating individual soil chemical and microbial variables, but mainly in terms of "matches" between them. To accomplish this we ran a series of standard (PERMANOVA) and not standard (BGA and Procrustes associate with ANOVA) statistical approaches considering different sampling origins.

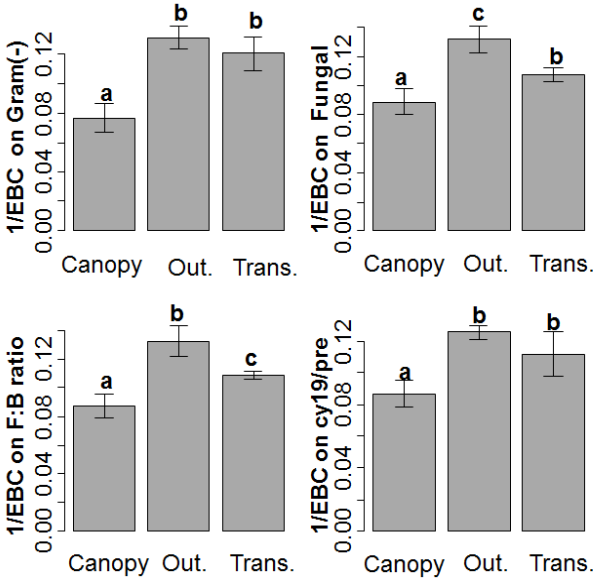


Figure 7. Mean test (LSD at 5% of significance; *P* value adjusted by Bonferroni) showing the variation in the match size between EBC and soil microbial variables and on microbial groups and indices as affected by the soil sample origin within the integrated crop-livestock-forest land use (iCLF). All these matches coming from partial PROTEST (Procrustes analyses) are based on Euclidian of log(x+1) dissimilarities matrices of individual variables. Canopy (samples from tree stand); Out (outside: samples from center of the pasture); Trans (transition: samples from canopy projection). F:B (Fungal:Bacterial ratio); cy19/pre (cy19/18:1ω7). The expression 1/match is used because these matches/effects are Procrustes association metrics representing the matching between individual soil chemical and soil microbial variables, which are inversely proportional to the level of relationship (or effect), so a higher value indicates a "lower effect". Standard errors (SE) are shown in each bar.

Table 3. One-way ANOVA showing which of the significant matches/effects of soil variables on microbial groups and indices (PLFA) given by the partial Protest from Table 2 were affected by the sample origin within integrated crop-livestock-forest land use type (iCLF)(sample origin factor levels: canopy, transition, outside).

	Sample origin within (iCLF)	
Relationships(matches)	*F*	*P*
EBC x Gram(-)	8.618	0.0047**
EBC x Fungal	7.292	0.0084**
EBC x F:B ratio	7.707	0.0070**
EBC x cy19/18:1ω7	4.005	0.0400*

*$P < 0.05$; **$P < 0.01$. "EBC" is the composite variable: exchangable base cations (Ca^{2+}, Mg^{2+} and K^{+}). F:B (Fungal:Bacterial ratio).

4.6.1. Sample position within the iCLF system affects microbial-based land use variations

Soil resource heterogeneity is well known as a biological diversification factor (Hodge, 2006), whereas variations in landscape characteristics due to the introduction of plant species are related to the heterogenization of different ecosystem components, including the soil microbial community (Bach et al., 2010; Carson et al., 2010). Thus, using pasture together with exotic forest species may make the iCLF system a consistent management strategy for seeking diversification of the quality of substrates offered to the soil microbial community (Vallejo et al., 2012). It is recognized that although lipid analyses have low taxonomic resolution compared to modern molecular techniques, the detailing of the microbial community into its main functional members, including fungi and bacteria, is sufficient to demonstrate that different land-use types also differ from each other in their microbial makeup (Lacombe et al., 2009; Unger et al., 2012; Vallejo et al., 2012).

Some studies have suggested that introducing tree species into pastures may increase the soil microbial diversity. Carson et al. (2010), for example, found that introducing *Eucalyptus* species into pastures favored the maintenance of soil fungal community diversity. Similarly, (Lacombe et al., 2009) reported that microbial community stability, associated with its heterogeneity, was higher in tree-based systems. The results obtained in this study through the strategy of sequentially using samples from different sampling origins within the iCLF in the BGA corroborate these previous findings. In our results fungal and bacterial variables were responsible for the main contrast between land use types in terms of microbial community structure, with iCLF and improved pasture both exhibiting a high fungal dominance trend when compared to degraded pasture and native forest. However it is not in line with the findings which observed fungal dominance in semi improved and unimproved pasture (Grayston et al., 2004).

Bacteria, fungi, and the F:B ratio were able to discriminate land use types regardless of sample position within the iCLF system. This supports other work showing that these two groups are suitable for discriminating land use types (Strickland and Rousk, 2010). In this point it is noteworthy that the differentiation between land use types based on the Gram (-) bacterial profile and the AMF marker (16:1ω5c) was affected by sample position within the iCLF. For example, the AMF marker was unable to discriminate land use when samples from the center of the pasture within the iCLF system (outside) were considered, but was successful when the samples from the forest component of iCLF (canopy) were used. These results highlighted the sensitivity of AMF since in the iCLF system the pasture component undergoes recurring interventions such as mechanization and animal introductions, which are factors that negatively affect the AMF (Jansa et al., 2002; van Groenigen et al., 2010),

whereas the tree stands within the forest component of the iCLF are not subject to the same management intensity.

The same argument used above may be used to try to explain why the land use types were discriminated by the Gram (-) bacterial profile when the samples from the center of the pasture component (outside) were considered rather than those from the tree stand (canopy). It may be due to the fact that the predominance of Gram (-) bacteria is attributed to conditions of higher resource availability (Ponder Jr and Tadros, 2002), which are found in the pasture component of the iCLF system because fertilizer addition is a practice used when the iCLF is in the grain production phase, affecting nutrient availability in the following pasture phase as well (Bardgett et al., 2001). These results are in line with Grayston et al. (2004), which indicated differences for different levels of management quality by observing that Gram (-) bacteria were dominant within improved pasture. However, it does not necessarily mean that all iCLF is dominated by Gram (-) bacteria at all, given that our results have also indicated a fungal dominance trend within this particular land use type. It suggests that the differences in sample origin are important determinants in differentiating the iCLF from other land use types.

Interestingly our results show that soil organic matter (SOM) was unable to differentiate the treatments, i.e., land use types (Table 1). Additionally, both the PERMANOVA analyses conducted with the canopy samples and those that used the soil from the center of the pasture (outside) in the iCLF system did not indicate SOM as a significant discriminator of land use types. It is likely due to the fact that SOM has been weakly related to other soil variables (Fig. 2abc). Also, the PCA results based on soil variables has highlighted that SOM had no significant contribution for land use type ordination regardless in the scenario of sampling origin (Fig. 3abc). Thus these results indicate that sample origin in iCLF is not able to discriminate from other land use types in terms of SOM, and also that the introduction of exotic tree species in the pasture has still not been able to generate contrasts in SOM within the iCLF. It corroborates results from Lai et al. (2007) who found little organic carbon variation in response to introduction of tree species into pastures. In contrast to organic matter, our results showed that the land use types were separated in all of the PERMANOVA contexts (all samples, canopy, outside) by moisture content, bulk density, and exchangeable base cations (Ca^{2+}+, Mg^{2+}, K^{+}), which suggests that the main difference of land use are associated with changes in these soil variables (Berthrong, 2009; Drenovsky et al., 2004). Moreover, the PCA results based on soil variables showed these variables as significantly related to ordination axes contributing towards highlighting the importance of these variables. For example, the bulk density was found to be higher in iCLF and degraded pasture (DP) than in improved pasture (IP) and native fragment (NF). Studies have indicated the influence of management intensity on the bulk density, by pointing out that areas not undergoing frequent anthropic interferences tend to show lowest bulk density values as result of the accumulation of plant residues incorporated into the soils, associated with non-disturbance of structure by machines, agricultural traffic and animal trampling (Hamza and Anderson, 2005). It may explains the low bulk density in the native fragment NF and improved pasture IP since while the former is an area without any anthropic interference, the latter is subjected to a low level of mechanization and grazing pressure when compared to iCLF and DP.

The accumulation of plant residues along with the absence of management pressure can also explain the relatively higher exchangeable base cations in the native fragment (NF) than other land use types. Also it is interesting to note that in our results the improved pasture (IP) remained similar to iCLF in terms of low exchangeable base cations availability across the two first scenarios of sample origin (all samples and canopy samples). However these two land use types became more divergent when samples from pasture within iCLF were considered (Fig. 3abc). These results suggest that the spatial variation in the exchangeable

base cations availability within the iCLF is driven by the man-generated plant heterogeneity characterizing this land use type, as well as it influencing how the iCLF differs from the improved pasture. Thus it seems that the area under tree influence has been responsible for making the iCLF a bit closer to improved pasture (IP) in terms of exchangeable base cations than the area under most pasture influence.

4.6.2. Microbial structure – soil properties "match size" variation

The individual effects of the soil variables om individual soil microbial structure variables which normally changed in response to alterations in land use type were indicated by two analogous analyses: the partial Mantel test and partial PROTEST (Procrustes analysis). The fact that the partial PROTEST provided a higher number of significant relationships between soil and microbial variables attests to the high power of this analysis approach compared to the Mantel test (Peres-Neto and Jackson, 2001).

As indicated by Procrustes analysis the variables soil bulk density, P availability, moisture content and exchangeable base cations showed a significant match to the microbial structure variables (Table 2). This reinforces the hypothesis that these variables were important in linking the changes in land use with the microbial community at our study site. However, only few of these matches/effects were affected by land use type (Table 2). For example, the matches between the composite variable exchangeable base cations (EBC) and microbial variables, namely Gram (-) and cy19/18:1ω7, were significantly affected by the land use type. This suggests that the EBC may be important, linking changes above and below ground. These results do agree with the increasing evidence of the important role of exchangeable base cations as drivers of the microbial community. For example, Allison et al., (2007) studying temporal and soil depth effects on microbial community structure found that SOM was not the main driver of microbial community composition, in contrast to the EBC.

Minerals are the primary source of EBC in the soils and it has been argued that the variation in the distributions of minerals in the soil may influence soil microbial variations (Carsson et al., 2009, Reith et al., 2012). For example, Gleeson et al. (2005,2006a) showed that singular bacterial and fungal communities colonized different minerals. However, despite these interesting previous findings it is unlikely that the mineral variability is the main mechanism explaining the partitioning of the matches between the EBC and the microbial variables by the land use type. The reason for this is that the entire area encompassing all land use types investigated has the same geological formation, being characterized by an intense soil weathering, which in turn is related to a narrow clay mineral range, namely kaolinite and iron oxides. Since this scenario stands for low natural soil fertility, it is more likely that EBC-microbial variable matches differences across land use types are due to the management history rather than geological formations. Our results showed that the EBC are high in the native fragment (NF) and low in the other land use types (Fig. 3abc). Interestingly the NF exhibited the weaker matches between EBC and microbial variables related to fertility status as Gram (-) and cy19/18:1ω7 (Ponder and Trados, 2002; Aliasgharzad et al., 2010) when compared to the other land use types, specially DP and iCLF (Fig. 5). It suggests low starvation effects on Gram (-) community in the NF whereas in the other man-managed land use types, specially DP and iCLF, the bacterial community seems to be more affected by the lack of resources as a result of lower EBC exhibited.

The PCA on soil chemical variables and the correlations among individual soil chemical variables showing that pH was positively linked to exchangeable base cations supports low matches of pH to microbial structure variables in DP and iCLF. This may be due to the high soil acidity and thus to the low nutrient availability. The relative distinctiveness of fungi and bacteria in relation to acidic environmental preferences has been well documented,

with fungal communities tending to dominate in the more acidic soils than bacteria (Rousk et al., 2009; Strickland and Rousk, 2010). We found weak matches between pH and microbial structure variables in DP and iCLF (Fig. 5), and within these matches the response of F:B ratio was a general measure of microbial shifting structure (Strickland and Rousk, 2010). Thus one would be expecting fungi to be dominant in DP and iCLF rather than in NF and iCLF. Our results indicated a fungal dominance trend in iCLF followed by IP (Fig. 4). Thus it is likely that the acidity in the DP is not associated with fungal dominance trend but rather with a bacterial community more adapted to acidity and starvation conditions as Gram (+), which was partially supported by our results (Fig. 4ab).

The variation in P availability had significant matches with microbial structure variables and supports the importance of this nutrient as a driver of soil microbial community, especially in tropical conditions (Liu et al 2012, 2013). Additionally the matches/effects size between P and microbial structure variables were observed to be more partitioned by land use type than the matches of other soil properties (Table 2). Interestingly, the effects of P on microbial variables such as F:B ratio, were less intense (weaker matches) in the iCLF than in the other treatments (Fig. 6). Although it is known that P affects the microbial community (Liu et al., 2012, 2013; Zhang et al., 2013), it is striking that these effects on the microbial community were less prominent in a unique land use type, namely iCLF, even though P availability was also low for other sites (Fig. 3). By associating these results with those reporting that iCLF has low P, low EBC (Fig. 3) and fungal dominance trend (Fig. 4), it is suggested that that microbial community found at iCLF is adapted to low resource availability (Rinnan and Baath, 2009; Rousk and Baath, 2007). We recognize, however, that resource availability is usually related to other important soil nutrients, especially N. Even so, these results highlight the importance of the P in our study area, which is characterized by high soil immobilization of this element.

4.6.3. The match between exchangeable base cation and soil microbial variables within the iCLF

Only the effects (matches) of EBC on microbial structure variables were partitioned by the sampling position within the iCLF. Clearly, the canopy and outside positions differed regarding magnitude of these effects (Fig. 7), with F:B ratio, Fungi, cy19/18:1ω7c and Gram (-) being higher under the canopy. As stated the canopy represents the area under highest influence of tree plantation whereas outside is the area the under low influence of the tree component. We have discussed in previous sections that the outside samples within the iCLF made this land use type more distinct from improved pasture in terms of EBC availability whereas the canopy samples made these two a bit closer as showed by PCA results (Fig. 3abc). It indicates that the canopy and outside samples within iCLF differ in EBC suggesting that the man-generated plant heterogeneity within iCLF may be shifting the soil microbial community structure by changing their response to base cations (Bach et al., 2010; Carson et al., 2010; Diedhiou et al., 2009).

4.7. CONCLUSIONS

It is mandatory to stress that despite interesting insights raised by this study, any temporal and spatial evaluations were not carried out. More studies are needed to address spatial and temporal consistencies of the land use type effects on the matches between microbial community and soil properties. Even so, the use of the Procrustean metric associated with downstream statistical approaches may open an interesting avenue for soil microbial ecologists, as it allows them to put the correlation as a central object of study either as response or predictor. In the present case by using Procrustes association metric in the ANOVA framework, we were able to show that the land use type distinction can be driven not just by individual soil chemical and microbial variables. The partitioning of the match sizes between soil chemical and microbial variables across land use types was useful in showing that iCLF appears to be an alternative for sustainable management as it showed a fungal dominance trend in an environment with low pH and fertility. Furthermore, we gained insights that both P and EBC are the most important soil chemical variables linking changes above and below ground. However, while the responses of microbial variables to the P are more land use type dependent, the effects of EBC (Ca^{+2}, Mg^{+2}, K^{+}) on microbial community variables are mainly affected by the samples position. Finally, we concluded that increases in the heterogeneity of vegetation within integrated crop, pasture and forestry systems are an important driver of microbial community response to environmental changes, and may be one means by which to increase the sustainability of tropical agroecosystems.

4.8. REFERENCES

ALIASGHARZAD, N.; MARTENSSON, L. M.; OLSSON, P. A. Acidification of a sandy grassland favours bacteria and disfavours fungal saprotrophs as estimated by fatty acid profiling. **Soil Biology & Biochemistry**, v. 42, p. 1058-1064, 2010

ANDERSON, M. J. A new method for non-parametric multivariate analysis of variance. **Austral Ecology**, v. 26, p. 32-46, 2001

ALLISON, V.J.; YERMAKOV, Z.; MILLER, R. M.; JASTROW, J. D.; MATAMALA, R. Using landscape and depth gradients to decouple the impact of correlated environmental variables on soil microbial community composition. **Soil Biology & Biochemistry**, v. 39, p. 505-516, 2007

BACH, L. H.; GRYTNES, J.-A.; HALVORSEN, R.; OHLSON, M. Tree influence on soil microbial community structure. **Soil Biology & Biochemistry**, v. 42, p. 1934-1943, 2010

BARDGETT, R. D.; MCALISTER, E. The measurement of soil fungal: bacterial biomass ratios as an indicator of ecosystem self-regulation in temperate meadow grasslands. **Biology and Fertility of Soils**, v. 29, p. 282-290, 1999

BARDGETT, R. D.; JONES, A. C.; JONES, D. L.; KEMMITT, S. J.; COOK, R.; HOBBS, P. J. Soil microbial community patterns related to the history and intensity of grazing in sub-montane ecosystems. **Soil Biology & Biochemistry**, v. 33, 1p. 653-1664, 2001

BERTHRONG, S. T., JOBBAGY, E. G.; JACKSON, R. B.A global meta-analysis of soil exchangeable cations, pH, carbon, and itrogen with afforestation.**Ecological Applications**, v.19, 2p. 228–2241, 2009a

BRASIL. Projeto Radambrasil. Divisão de Publicação, Rio de Janeiro, 1983

CAO, Y.; FU, S.; ZOU, X.; CAO, H.; SHAO, Y.; ZHOU, L. Soil microbial community composition under Eucalyptus plantations of different age in subtropical China. **European Journal of Soil Biology**, v. 46, p. 128-135, 2010

CARSON, J. K.; CAMPBELL, L.; ROONEY, D.; CLIPSON, N.; GLEESON, D. B. Minerals in soil select distinct bacterial communities in their microhabitats. **FEMS Microbiology Ecology**, v. 67, p. 381-388, 2009

CARSON, J. K.; GLEESON, D. B.; CLIPSON, N.; MURPHY, D. V. Afforestation alters community structure of soil fungi. **Fungal Biology**, v. 114, p 580-584, 2010

CHESSEL, D.; DUFOUR, A. B.; THIOULOUSE, J. The ade4 package-I-One-table methods. **R news**, v. 4, 5-10, 2004

DIEDHIOU, S.; DOSSA, E. L.; BADIANE, A. N.; DIEDHIOU, I.; SÈNE, M.; DICK, R. P. Decomposition and spatial microbial heterogeneity associated with native shrubs in soils of agroecosystems in semi-arid Senegal. **Pedobiologia**, v. 52, p. 273-286, 2009

DON, A.; SCHUMACHER, J.; FREIBAUER, A. Impact of tropical land-use change on soil organic carbon stocks—a meta-analysis. **Global Change Biology**, v. 17, p. 1658-1670, 2011

DRENOVSKY, R. E.; VO, D.; GRAHAM, K. J.; SCOW, K. M. Soil water content and organic carbon availability are major determinants of soil microbial community composition. **Microbial Ecology**, v. 48, p. 424-430, 2004

EMBRAPA. Manual de análises químicas de solos, plantas e fertilizantes [Manual of chemical analyses of soil, plants, and fertilizers], 2° ed. Embrapa Informação Tecnológica, Brasília, 2009

FERNANDES, M. F.; BARRETO, A. C.; MENDES, I. C.; DICK, R. P. Short-term response of physical and chemical aspects of soil quality of a kaolinitic Kandiudalfs to agricultural practices and its association with microbiological variables. **Agriculture, Ecosystems & Environment**, v. 142, p. 419-427, 2011

FROSTEGÅRD, Å.; BÅÅTH, E.; TUNLID, A. Shifts in the structure of soil microbial communities in limed forests as revealed by phospholipid fatty acid analysis. **Soil Biology & Biochemistry**, v. 25, p. 723-730, 1993

FROSTEGÅRD, Å., TUNLID, A., BÅÅTH, E. Use and misuse of PLFA measurements in soils. **Soil Biology & Biochemistry**, v. 43, p. 1621-1625, 2011

GIRALDO, C.; ESCOBAR, F.; CHARÁ, J. D.; CALLE, Z. The adoption of silvopastoral systems promotes the recovery of ecological processes regulated by dung beetles in the Colombian Andes. **Insect Conservation and Diversity**, v. 4, p. 115-122, 2011

GOWER, J. C. Statistical methods of comparing different multivariate analyses on the same data, In: HODSON, F.R.; KENDALL, D. G.; TAUTU, P. (Eds.), Mathematics in the Archeological and Historical Sciences. Edinburgh University Press, Edinburgh, pp. 138-149, 1971

GRAYSTON, S.J., CAMPBELL, C.D., BARDGETT, R.D., MAWDSLEY, J.L., CLEGG, C.D., RITZ, K., GRIFFITHS, B.S., RODWELL, J.S., EDWARDS, S.J., DAVIES, W.J., ELSTON, D.J., MILLARD, P. Assessing shifts in microbial community structure across a range of grasslands of differing management intensity using CLPP, PLFA and community DNA techniques. **Applied Soil Ecology**, v. 25, p. 63-84, 2004

GLEESON, D.; CLIPSON, N.; MELVILLE, K.; GADD, G.; MCDERMOTT, F. Characterization of fungal community structure on a weathered pegmatitic granite. **Microbiology Ecology**, v. 50, p. 360–368, 2005

GLEESON, D.; KENNEDY, N.; CLIPSON, N.; MELVILLE, K.; GADD, G.; MCDERMOTT, F. Characterization of bacterial community structure on a weathered pegmatitic granite. **Microbiology Ecology**, v. 5, p. 526–534, 2006a

HAMZA, M. A.; ANDERSON, W. K. Soil compaction in cropping systems: A review of the nature, causes and possible solutions. **Soil and Tillage Research**, v. 82, p. 121-145, 2005

HODGE, A. Plastic plants and patchy soils. **Journal of Experimental Botany**, v. 57, p. 401-411, 2006

JANSA, J.; MOZAFAR, A.; ANKEN, T.; RUH, R.; SANDERS, I.; FROSSARD, E. Diversity and structure of AMF communities as affected by tillage in a temperate soil. **Mycorrhiza**, v. 12, p. 225-234, 2002

KASEL, S.; BENNETT, L.T.; TIBBITS, J. Land use influences soil fungal community composition across central Victoria, south-eastern Australia. **Soil Biology & Biochemistry**, v. 40, p. 1724-1732, 2008

KUO, S. Phosphorus, In: Sparks, D.L. (Ed.), Methods of Soil Analysis: Chemical Methods (Part 3). SSSA, Madison, WI, pp. 869-919, 1996

LACOMBE, S.; BRADLEY, R. L.; HAMEL, C.; BEAULIEU, C. Do tree-based intercropping systems increase the diversity and stability of soil microbial communities? **Agriculture, Ecosystems & Environment**, v.131, p. 25-31, 2009

LAI, R.; LAGOMARSINO, A.; LEDDA, L.; ROGGERO, P. P.Variation in soil C and microbial functions across tree canopy projection and open grassland microenvironments. Turkish **Journal of Agriculture and Forestry**, v. 38, p. 62-69, 2014

LAMB, E. G.; KENNEDY, N.; SICILIANO, S. D. Effects of plant species richness and evenness on soil microbial community diversity and function. **Plant and Soil**, v. 338, p. 483-495, 2011

LEGENDRE, P.; LEGENDRE, E.Numerical Ecology, 3rd English edn. Elsevier344 Science BV, Amsterdam. 516 p, 2012

LISBOA, F. J. G.; CHAER, G. M.; JESUS, E.D. C.; FARIA, S. M.; GONÇALVES, F. S.; SANTOS, F. M.; CASTILHO, A. F.; BERBARA, R. L. L. The influence of litter quality on the relationship between vegetation and below-ground compartments: a Procrustean approach. **Plant and Soil**, v. 367, p. 551-562, 2012

LISBOA, F. J. G.; PERES-NETO, P. R.; CHAER, G. M.; JESUS, E. C.; MITCHELL, R. J.; CHAPMAN, S. J.; BERBARA, R. L. L. Much beyond Mantel: bringing Procrustes Association Metric to the plant and soil ecologist's toolbox. **PLoS One**, v. 9, e101238, 2014

LIU, L.; GUNDERSEN, P.; ZHANG, T.; MO, J. Effects of phosphorus addition on soil microbial biomass and community composition in three forest types in tropical China. **Soil Biology & Biochemistry**, v. 44, p. 31-38, 2012

LIU, L.; ZHANG, T.; GILLIAM, F. S.; GUNDERSEN, P.; ZHANG, W.; CHEN, H.; MO, J. Interactive effects of nitrogen and phosphorus on soil microbial communities in a tropical forest. **PLoS One**, v. 8, e61188, 2013

MENDIBURU, F. agricolae: Statistical Procedures for Agricultural Research. 2014 http://cran.r-project.org/web/packages/agricolae/index.html.

MILLARD, P.; SINGH, B. K. Does grassland vegetation drive soil microbial diversity? **Nutrient Cycling in Agroecosystems**, v. 88, p. 147-158, 2009

MITCHELL, R. J.; HESTER, A. J.; CAMPBELL, C. D.; CHAPMAN, S. J.; CAMERON, C. M.; HEWISON, R. L.; POTTS, J. M. Is vegetation composition or soil chemistry the best predictor of the soil microbial community? **Plant and Soil**, v. 333, p. 417-430, 2010

NELSON, D. W.; SOMMERS, L. E. Total carbon, organic carbon and organic matter, In: SPARKS, D.L., PAGE, A. L.; HELMKE, P. A.; LOEPPERT, R. H.; SOLTANPOUR, P. N.; TABATABAI, M. A.; JOHNSTON, C. T.; SUMNER, M. E. (Eds.), Methods of Soil Analysis: Chemical Methods (part 3). SSSA, Madison, WI, pp. 961-1010, 1996.

OKSANEN, J.; BLANCHET, F. G.; KINDT, R.; LEGENDRE, P.; MINCHIN, P. R.; O'HARA, R. B.; SIMPSON, G. L.; SOLYMOS, P.; STEVENS, M. H. H.; WAGNER, H. vegan: Community Ecology Package, 2013 http://cran.r-project.org/web/packages/vegan/index.html.

OLSSON, P. A. Signature fatty acids provide tools for determination of the distribution and interactions of mycorrhizal fungi in soil. **FEMS Microbiology Ecology**, v. 29, p. 303-310, 1999

PERES-NETO, P.; JACKSON, D. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. **Oecologia**, v. 129, p.169-178, 2001

PONDER JR, F., TADROS, M. Phospholipid fatty acids in forest soil four years after organic matter removal and soil compaction. **Applied Soil Ecology**, v. 19, p. 173-182, 2002

R Core Development Team, 2013. http://www.r-project.org/.

REITH, F.; BRUGGER, J.; ZAMMIT, C. M.; GREGG, A. L.; GOLDFARB, K. C.; ANDERSEN, G.L., DESANTIS, T. Z.; PICENO, Y. M.; BRODIE, E. L.; LU, Z.; HE, Z.; ZHOU, J.; WAKELIN, S. A. Influence of geogenic factors on microbial communities in metallogenic Australian soils. **ISME Journal** 6, 2107-2118, 2012

RINNAN, R.; BAATH, E. Differential utilization of carbon substrates by bacteria and fungi in tundra soil. **Applied and Environmental Microbiology**, v. 75, p. 3611-3620, 2009

RODRIGUES, J. L.; PELLIZARI, V.H.; MUELLER, R.; BAEK, K.; JESUS, EDA, C.; PAULA, F. S.; MIRZA, B.; HAMAOUI, G. S. JR.; TSAI, S. M.; FEIGL, B.; TIEDJE, J. M.; BOHANNAN, B. J.; NUSSLEIN, K. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proceedings of the National Academy of Sciences**, v. 110, p. 988-993, 2013

ROUSK, J.; BAATH, E. Fungal and bacterial growth in soil with plant materials of different C/N ratios. **FEMS Microbiology Ecology**, v. 62, p. 258-267, 2007

ROUSK, J.; BROOKES, P. C.; BAATH, E. Contrasting soil pH effects on fungal and bacterial growth suggests functional redundancy in carbon mineralisation. **Applied and Environmental Microbiology**, v.75, p. 1589–1596, 2009

STRICKLAND, M. S.; ROUSK, J. Considering fungal:bacterial dominance in soils—Methods, controls, and ecosystem implications. **Soil Biology & Biochemistry**, v. 42, p. 1385-1395, 2010

THIOULOUSE, J.; PRIN, Y.; DUPONNOIS, R. Multivariate analyses in soil microbial ecology: a new paradigm. **Environmental and Ecological Statistics**, v.19, p. 499-520, 2012

THOMAS, G. W. Soil pH and soil acidity, In: Sparks, D.L. (Ed.), Methods of Soil Analysis: Chemical Methods (Part 3). SSSA, Madison, WI, pp. 475-490, 1996

TUOMISTO, H.; RUOKOLAINEN, K.; AGUILAR, M.; SARMIENTO.; A. Floristic patterns 354 along 43-km long transect in an Amazonian rain forest. **Journal of Ecology**, v. 91, p. 743-756, 2003

UNGER, I. M.; GOYNE, K. W.; KREMER, R. J.; KENNEDY, A. C. Microbial community diversity in agroforestry and grass vegetative filter strips. **Agroforestry Systems**, v. 87, p. 395-402, 2012

VALLEJO, V.E., ARBELI, Z., TERÁN, W., LORENZ, N., DICK, R.P., ROLDAN, F. Effect of land management and Prosopis juliflora (Sw.) DC trees on soil microbial community and enzymatic activities in intensive silvopastoral systems of Colombia. **Agriculture, Ecosystems & Environment**, v. 150, p. 139-148, 2012

VAN GROENIGEN, K.-J.; BLOEM, J.; BÅÅTH, E.; BOECKX, P.; ROUSK, J.; BODÉ, S.; FORRISTAL, D.; JONES, M. B. Abundance, production and stabilization of microbial biomass under conventional and reduced tillage. **Soil Biology & Biochemistry**, v. 42, p. 48-55, 2010

WIXON, D. L.; BALSER, T.C. Toward conceptual clarity: PLFA in warmed soils. **Soil Biology & Biochemistry**, v. 57, p. 769-774, 2013

WRIGHT, R.J.; STUCZYNSKI, T. Atomic absorption and flame emission spectrometry, In: SPARKS, D.L. (Ed.), Methods of Soil Analysis: Chemical Methods (Part 3). SSSA, Madison, WI, pp. 65-90, 1996

ZELLES, L. Fatty acid patterns of phospholipids and lipopolysaccharides in the characterization of microbial communities in soil: a review. **Biology and Fertility of Soils**, v. 29, p. 111-129, 1999

ZHANG, X.; CHEN, Q.; HAN, X. Soil bacterial communities respond to mowing and nutrient addition in a steppe ecosystem. **PLoS One** v. 8, e84210, 2013

5. CAPÍTULO III:

EXPLORING THE EFFECTS OF THE CORRELATION WITHIN AND BETWEEN DATA TABLES ON PROCRUSTES ANALYSIS OUTPUT: DIRECTIONS FOR PLANT AND SOIL ECOLOGISTS

Capítulo submetido como: Lisboa, F. J. G.; Mitchell, R. J.; Chapman, S.; Potts, J.M.; Berbara, R. L. L. Exploring the effects of the correlation within and between data tables on Procrustes analysis output. **PloS One**, v. x, p. xx-xx, 2015

5.1. RESUMO

O vetor de resíduos Procrusteanos (ou PAM, um acrônimo para o termo equivalente em inglês: *Procrustean association metric*) derivado da análise Procrustes pode ser visto como a forma univariada do relacionamento entre duas ou mais tabelas de dados multivariados, o que fornece interessante maneira para que ecologistas coloquem o relacionamento multivariado como objeto central de investigações em abordagens estatísticas mais familiares, como ANOVA e comparação de médias. Porém, muitos aspectos precisam sem melhor elucidados no sentido de tornar ecologistas mais confidentes em usar a Procrustes em seus estudos. Aqui, foram exploradas duas questões comumente levantadas por ecologistas não versados na Procrustes: 1) Usando dados do segundo capítulo da presente tese (PLFA: tabela **Y**; Fertilidade: tabela **X**), foi indagado: **i)** como o crescente número de colunas/variáveis correlacionadas em diferentes níveis dentro de uma das tabelas de dado usadas na Procrustes (tabela **X**) é capaz de afeta os resultados da análise? **ii**) Usando tabelas de dados simuladas perguntamos: a PAM é capaz de mostrar como a correlação multivariada entre tabelas de dados é particionada por diferentes tratamentos? Observou-se que o crescente número de variáveis correlacionadas (6, 9, 12 e 15) em diferentes níveis (0,9, 0,7, 0,5 e 0,2), dentro da tabela de dados não sujeita às transformações inerentes à análise Procrustes (translação e rotação), no presente caso a tabela **X**, não teve efeito sobre resultados clássicos da Procrustes relacionados ao seu ajuste (estatística *R* e sua significância). Também, o crescente número de variáveis correlacionadas em diferentes níveis não teve claro efeito sobre a significância da ANOVA usando a PAM como resposta. Por outro lado, a crescente correlação entre duas tabelas de dados, **X** e **Y**, para uma parte específica das mesmas representando hipotético tratamento foi detectada pela PAM quando esta foi usada em comparação múltipla de médias. Coletivamente nossos resultados suportam que as estatísticas Procrusteanas levam apenas em consideração a informação existente entre as tabelas de dados sob investigação, e que a PAM de fato reflete a diferença em termos de correlação multivariada quando esta é usada em procedimentos estatísticos mais tradicionais, como a comparação de médias, o que pode ser útil para ecologistas de planta e solo, e ecologistas de uma forma geral, interessados em levantar indícios sobre a variação da correlação multivariada entre diferentes níveis categóricos (parcelas, paisagens, tipos de uso da terra, gradientes ambientais, etc.).

Palavras-chave: Análises multivariadas. Métrica de associação. Tabela de dados

5.2. ABSTRACT

The Procrustean residual vector (or PAM, an acronym for the alternative equivalent term Procrustean association metric) derived from Procrustes analysis can be seen as the univariate form for the relationship between two or more data tables, which provides an interesting way for ecologists to place multivariate relationships as the central object of investigation in more familiar statistical approaches such as ANOVA and post hoc tests. However, many aspects need to be elucidated to become ecologists more confident on using the Procrustes in their studies. Here we attempted to explore two questions: **i**) How does the increasing number of correlated columns within an entire data table affect the Procrustes results? **ii**) Can the PAM be used for detecting how the correlation is partitioned across treatments levels within the original data table? We found that the increasing number of correlated variables (6, 9, 12, and 15) across different imposed correlation levels (0. 9, 0.7, 0. 5, and 0.2) in the data table not subject to Procrustean linear transformation (translation and rotation), i.e. the X data table, had no effects on classical Procrustes outcomes related to the fit between data tables(*R* statistic and its *P* value) and on the significance of the ANOVA using the Procrustes residual vector (PAM), which summarizes the multivariate correlation between two data tables, as response. On other hand increasing the between correlation levels between **X** and **Y** data tables for a specific treatment resulted in PAMs that, when used in mean multiple comparisons, did show this categorical level as different from all others. Collectively our results supports that Procrustes fit is only dependent of the information between data table instead of within data table, and the PAM reflects the differences in multivariate correlation across data tables when it is used in downstream statistical approach, such as multiple comparisons of means, what can be useful for ecological questions addressing the partitioning of the multivariate correlation among different categorical levels (plots, time, land use type, etc).

Keywords: Multivariate analysis. Association metric. Data tables

5.3. INTRODUCTION

Analysis of Variance (ANOVA) is used as a tool to split the variability of a given outcome of interest into basic components like: 1) the variability explained by one or more categorical predictors; 2) the residual variation. In the ANOVA framework the response variables can vary in their nature, being classified as continuous or discrete, and univariate or multivariate. The simplest univariate context of ANOVA, that is, one response and one categorical predictor is obviously easier to analyze than the multivariate context; however for ecologists the univariate world rarely exists given that many ecological questions require one to handle multiple variables. Therefore the question arises: how can ecologists fit the natural multivariate requirement of ecological research to the simplicity of the univariate ANOVA and post hoc test frameworks?

Lisboa et al. (2014a) showed how the results from Procrustes analysis (Gower, 1971),a multivariate statistical approach for correlating data tables representing sets of information coming from the same object of study: plots, Environmental gradient levels, experimental treatments, etc, could be used in downstream statistical analysis, including ANOVA and post hoc tests. Procrustes analysis has been shown to be statistically superior in some aspects (lower Type I error and higher power) than the traditional analogue approach, the Mantel test (Peres-Neto and Jackson, 2001)and one of the features that arise from Procrustes analysis is the possibility of providing the multivariate relationship among two, or more, data tables in a vector form make by residuals, the Procrustean residual vector, also named the Procrustean association metric (PAM) (Lisboa et al. 2014a). For example, assuming that an ecologist wants to correlate two data tables, **X** and **Y**, the first one representing abiotic variables (climate, soil, elevation, etc), and the second one representing a certain biological community (birds, bacterial, etc). Moreover let`s assume that in **X** and **Y** all variables (columns) were measure from fifth plots, which represent the rows of the data tables. Procrustes analysis will found the "best" fit of homologue coordinates across **X** and **Y** data tables by seeking to minimize of the sum of squares between correspondent coordinates in **X** and **Y**, i. e. the plots or rows of these tables. Given that the Procrustes fit is never perfect, the PAM stands for the residuals between corresponding coordinates (the plots or rows of X and Y) after the "best" fit among the tables has been found so that the lower residual sizes in the PAM, the higher the correlations. The compilation of these residual differences between homologous rows in the X and Y data tables making up the Procrustean residual vector (PAM) and represents a useful way to represent information on the relationship between two matrices and make it available for further statistical analysis, both parametric and non-parametric (Lisboa et al., 2014a).

Despite of the potential uses of this Procrustean feature in ecological research, the rigor of this composite framework (made up of PAM – ANOVA – Post hoc tests) has not been assessed. In particular two: 1) how does the increasing number of columns correlating within an entire data table affect the Procrustes results? 2) Can the PAM be used for detecting how the correlation is partitioned across treatments levels within the original data table? For simplicity, the Procrustes analysis results can be divided into two sets 1) **mainstream results**, which take account of the correlation statistic (*R*) between multidimensional data tables and its significance: (*P* value), i. e. the Procrustes fit; and 2) **downstream results**, which are related to statistics provided by the analysis of the PAM using other statistical frameworks, like ANOVA and multiple comparisons of means using, for example, Tukey's HSD test.

With respect to the first question, Dray et al (2003) argued that the Procrustes fit, that is, the mainstream results, is only influenced by the variation between matrices. It arises that the variation in the number of columns correlating within a data table should not influence the mainstream results such as *R* statistic and its *P* value. However, nothing is known about the

consequences of the number of columns correlating within a data table on the PAM. For example, it was never explored whether the increasing correlation within an entire data table is conveyed to the PAM and whether it affects outcome as those from using PAM as response in ANOVA framework.

On other hand there are no papers exploring explicitly the following statement: “the analysis of the PAM, by looking at consistencies of the residual size between homologues coordinates across different treatments, could be useful for providing insights about differences in terms of multivariate correlation”. Such statement is linked to the use of the PAM in downstream statistical analysis such as multiple comparison of mean, for instance Tukey’s HSD test. Thus we investigated whether the PAM is able to detect differences in multivariate correlations among treatments when it is using in multiple comparison of means.

5.4. MATERIAL AND METHODS

5.4.1. First question: does the increasing number of correlated columns within an entire data table affect Procrustes results?

For assessing whether the increasing number of columns correlating within an entire data table affects the Procrustes results we used two data tables from a study by Lisboa et al. (2014b). In this study the authors used Procrustes analysis together with ANOVA and mean multiple comparisons in order to assess how the strength of the "match" (correlation/correlations) between individual soil microbial community variables and individual soil fertility variables varied across four land use types (native forest, degraded pasture, improved crop, integration crop-livestock-forest). The soil microbial community (PLFA profile) and the soil fertility data tables had the following dimensions: (n = 53, p = 20) and (n = 53, p = 15), respectively. Hereafter n and p stand for row and column numbers in the data tables, respectively.

Four correlation levels (0.2; 0.5; 0.7; and 0.9) were incorporated into different number of columns within the soil fertility data table (n = 53, p = 15), hereafter the **X** data table. The number of soil variables that were correlated was increased gradually (6, 9, 12, and 15) (see section 2.1.1 for method), whereas the PLFA profile data table (n = 53, p = 20), here **Y,** had its original correlation structure unaltered. After that **X** (soil fertility) and **Y** (PLFA profile) data tables were submitted to thirteen pre-transformations. These 13 pre-Procrustes transformations were used to encompass the three different manners in which **X** and **Y** data tables can be used in the Procrustes analysis: raw data, distance matrices, ordination axes (Fig. 1c)Finally the Procrustes relationships between **X** (soil fertility) and **Y** (PLFA profile) were simulated hundred times for each pre-transformation. Therefore from each set of 100 simulations / pre-transformation within a given number of correlated columns within X (6, 9, 12, 15) with a given correlation level (0.2; 0.5; 0.7; 0.9), we retained the following statistics: 1) the average of the Procrustean correlation statistic (R value); 2) the number of times that the R statistic was significant (P value); 3) the PAM average (a Procrustean residual vector from the average of the 100 simulations); 4) the residual size range within the PAM average (subtracting the highest and the lowest residual sizes linking the two data tabled after Procrustes fit); 5) the number of times in which the ANOVA using the PAM average as response and the land use type as factor (4 levels) was considered significant ($P < 0.05$). Thus, we retained 13 sets of Procrustes information which were used for making up graphs.

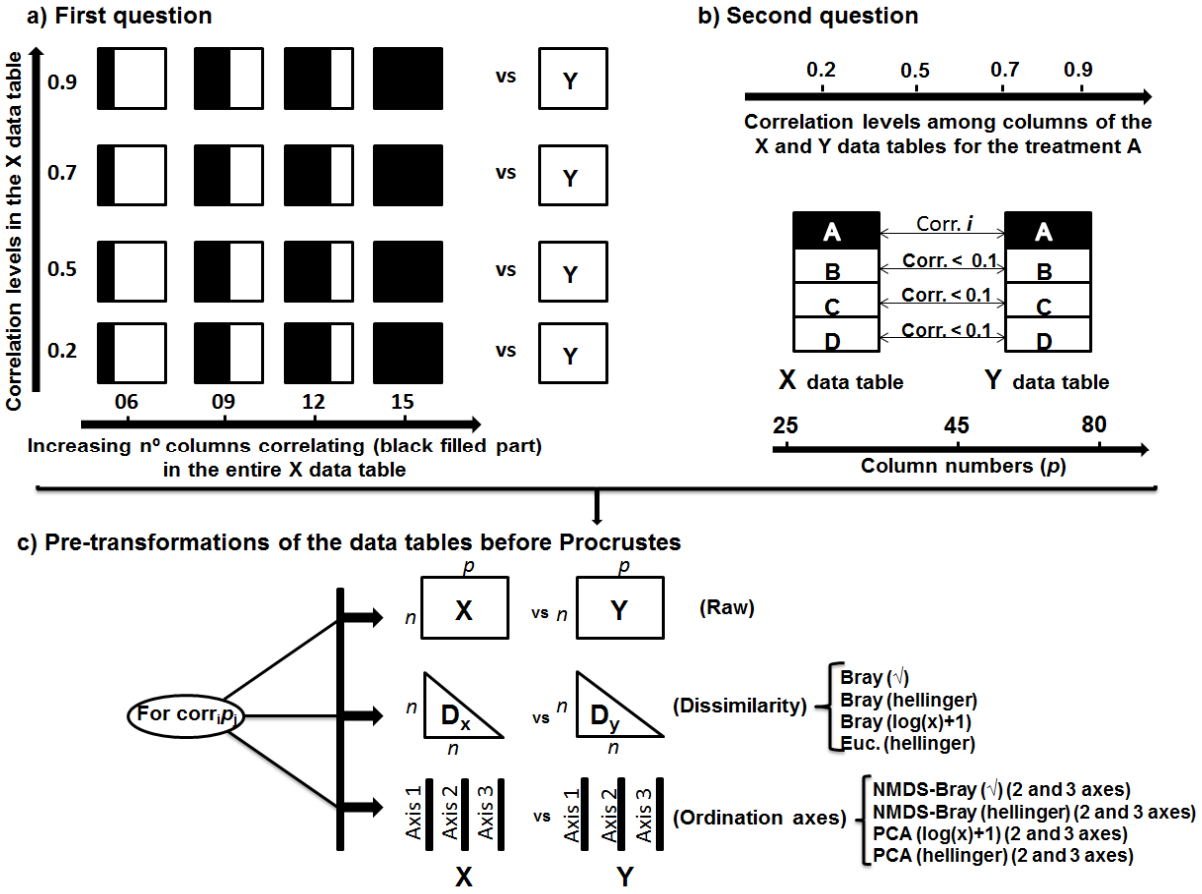


Figure 1. General approach used in the study. **a**) Illustration of the first question addressed: the effects of increasing the number of variables correlating within **X** data table (soil fertility, n = 53, p = 15) on the Procrustes results. For each number of columns (6, 9, 12, 15) in the **X** data table, four correlation levels were imposed (0.2; 0.5; 0.7, 0.9). The **X** data table was related to **Y** data table (PLFA profile, n =53, p = 20, none correlation structure imposed to it) by Procrustes analysis. **X** and **Y** data tables are from Lisboa et al. (2014b). **b**) Illustrates the second question: whether the correlation level between **X** and **Y** data tables (simulated data) incorporated into a specific treatment (treatment A) is reflected in the results of ANOVA analysis of the PAMs. The correlation between **X** and **Y** for all others treatments (B, C, and D) was not greater than 0.1. **c**) The different pre-Procrustes transformation in which **X** and **Y** were used in the Procrustes analysis (raw data, distance matrices, ordination axes). Each of these pre-transformations was simulated 100 times in order to get the results.

5.4.2. Correlation incorporated into soil fertility data table for accessing the first question

The process of incorporating distinct correlation levels into the soil fertility data table, the **X** data table, followed two basic steps:

a) Specific level - correlation matrix **M** generation (0.2, 0.5, 0.7, and 0.9);
b) Spectral decomposition of **M** *into* $\mathbf{LL}^{T}$, and multiplication of $\mathbf{L}^{T}$ by the transpose of the soil fertility matrix **X**.

For generating the specific-level correlation table **M**, we used the R functions described by Hardin at al. (2013) which are intended for building correlation matrices with

noise addition (http://pages.pomona.edu/~jsh04747/research/simcor.r.). Here the *noise* added to the **M** entries was from -0.001 to 0.001. After obtaining **M**,its correlation structure levels were incorporated into the soil fertility data table by using the following R code:

```
fert.unc<-t((solve(t(chol(cov(fert.m)))))%*%t(fert.m)) # makes X data table uncorrelated
object<- t(chol(M))%*% t(fert.unc) # incorporates correlation structure levels
objetc.df <- t(objetc)       # creates the simulated soil fertility data frame
corrplot(cor(object.df)) # checks the correlation structure incorporated
```

Specifically, all correlation structure levels (0.2; 0.5; 0.7; and 0.9) were incorporated into the entire **X** data table (n = 53, p =15) and from them the number of columns (variables) correlating was reduced gradually (15, 12, 9, and 6) within each correlation level. It means **X** data table having 15 (total), 12, 9, and 6 columns correlating at 0.2; 0.5; 0.7 and 0.9. The columns within the soil fertility that were left without any correlation level imposed. For example, for evaluating the effects of correlation levels imposed into 6 columns of the entire X data table (n = 53, p = 15) the rest of 9 columns within X were not correlated among them. It means that actually the soil fertility data table, the **X** data table, was only a template for our investigation and its original correlation does not matter. The whole process from incorporating different correlation levels into an increasing number of columns within the **X** data table to the Procrustes analysis was simulated 100 times for each one of the 13 pre-Procrustes transformations described in Fig. 1c.

5.4.3. Second question: can the PAM be used for detecting how the multivariate correlation is partitioned across different treatments?

For assessing whether the use of PAM in multiple comparisons of means is able to detect differences among treatments in terms of multivariate correlation, simulated data tables **X** and **Y** were used. One can visualized as if these data arose from a hypothetical scenario where **X** and **Y** are data tables derived from a study investigating how the multivariate correlation between general plant community (**X** data table) and its functional traits (**Y** data table) is partitioned across an environmental gradient based on the time elapsed after an intense burning event. Also, one can consider that **X** and **Y** are encompassing four times elapsed after burning event where plant community (**X** data table) and functional traits (**Y** data table) were measured times A, B, C, and D.

Hereafter these four times will be referred as treatments. Four different correlation levels (0.2; 0.5; 0.7; 0.9) were only incorporated into the treatment A for both **X** and **Y** data tables, and this treatment A corresponds to the first ten rows of each these data table. For all other treatments (B, C, and D) the correlation between **X** and **Y** was never greater than 0.1 (Fig. 1b). After correlation level incorporation, the 13 pre-Procrustes transformations were applied to both, **X** and **Y** data tables, before the Procrustes analysis (Fig. 1c). All steps from the **X** and **Y** data table generation to Procrustes analysis were repeated 100 times. Also, these simulations were carried out varying the number of columns (variables, p) in relation to the number of rows (sites, n) so that **X** and **Y** were data tables with the follow dimensions: (n = 40, p = 25); (n = 40, p = 45) and (n = 40, p = 80). From each set of 100 simulations / pre-transformation within a given correlation level between **X** and **B** data tables incorporated into to the treatment A (0.2; 0.5; 0.7; 0.9) we retained the following information from the Procrustes results: 1) the number of times in which the treatment A came out as being different from all other treatment (A ≠ B,C,D) when using the average PAM in Tukey HSD (95%); 2) the average number of the Procrustean residual size in each treatment (A, B, C, D). Thus, for each correlation level between **X** and **Y** in the treatment A, we retained 13 sets of Procrustes information which were used for making up graphs.

5.4.4. Different correlation levels between X and Y for a specific treatment

For creating the simulated data tables with different between correlation levels for a specific categorical level in both **X** and **Y** tables (namely level A) we first created sets of three "big" tables: 1 (n = 10, p = 50), 2 (n = 10, p = 90), 3 (n = 10, p = 160). Four correlation structure levels were incorporated into each "big" table (0.2; 0.5; 0.7; 0.9), and this was carried out using the same procedure described for the soil fertility data table in the first part of this paper.

After the correlation structure was added to the "big" data tables, each one was broken down into two equal tables. For example, in the case of a "big" table (n = 10, p = 50) with a given correlation level of 0.2, it was divided into $\mathbf{X}corr_{0.2}$ (n = 10, p = 25) and $\mathbf{Y}corr_{0.2}$ (n = 10, p = 25) tables. Thus, each one of these "big" tables provided four pairs of **X** and **Y** data tables (n = 10, p = 25, 45, 80) representing different correlation levels between them for the treatment A, such that: (A$\mathbf{X}corr_{0.9}$versus A$\mathbf{Y}corr_{0.9}$); (A$\mathbf{X}corr_{0.7}$versus AY$corr_{0.7}$); (A$\mathbf{X}corr_{0.5}$versus A$\mathbf{Y}corr_{0.5}$); (A$\mathbf{X}corr_{0.2}$ versus A$\mathbf{Y}corr_{0.2}$).

For taking account of other treatments (B, C, and D) we created three "big" tables: 1 (n = 30, p = 50), 2 (n = 30, p = 90), and 3 (n = 30, p = 160), with all columns p having the same correlation level (corr. < 0.1). These tables were broken down as in the same way as for treatment A. Thus, for each (n = 30, p = 25, 45, and 80) four pairs of **X** and **Y** tables were generated. The tables A$\mathbf{X}cov_i$ and A$\mathbf{Y}cov_i$ were then linked to BCD$\mathbf{X}corr_{<0.1}$ and BCD$\mathbf{Y}corr_{<0.1}$ tables, respectively, in order to build the entire **X** and **Y** tables (n = 40, p = 25, 45, 80) as shown in (Fig. 1b). The whole process of incorporating different correlation levels into **X** until the Procrustes analysis was simulated 100 times for each one of the 13 pre-Procrustes transformations described in Fig. 1c.

5.5. RESULTS

5.5.1. Imposed correlation effects on individual mainstream and downstream Procrustes results

Both classical Procrustes mainstream results, the Procrustean correlation statistic *R* and its significance (*P* value), remained constant irrespective of the increasing number of correlated variables within the **X** data table and the level of correlation incorporated into them (Fig. 2ab). The constancy across the increasing number of correlated columns and their imposed correlation levels within the **X** data table was also true for Procrustes results involving the PAM, such as the measure of residual size variability across individual PAMs, the residual ranging size (maximum minus minimum residual sizes in the PAM linking the two data tables under analysis), and the number of significant ANOVA results using PAMs as response variable (Fig. 2cd).

5.5.2 Raw data, distance matrices, and ordination generated similar Procrustes results

It was observed that the forms in which two data tables can be used in the Procrustes analysis (raw data, distance matrices, ordination axes) no clearly differed each other in relation the Procrustes outcomes regardless the imposed correlation level (Fig. 3). Moreover, it was observed that when the average PAMs for each pre-Procrustes treatment (100 simulations) were used in an ordination analysis (NMDS, "Euclidian distance"), the PAMs coming from different data type entries (raw data, distance matrices, ordination axes) were not divergent each other (Fig. 4).

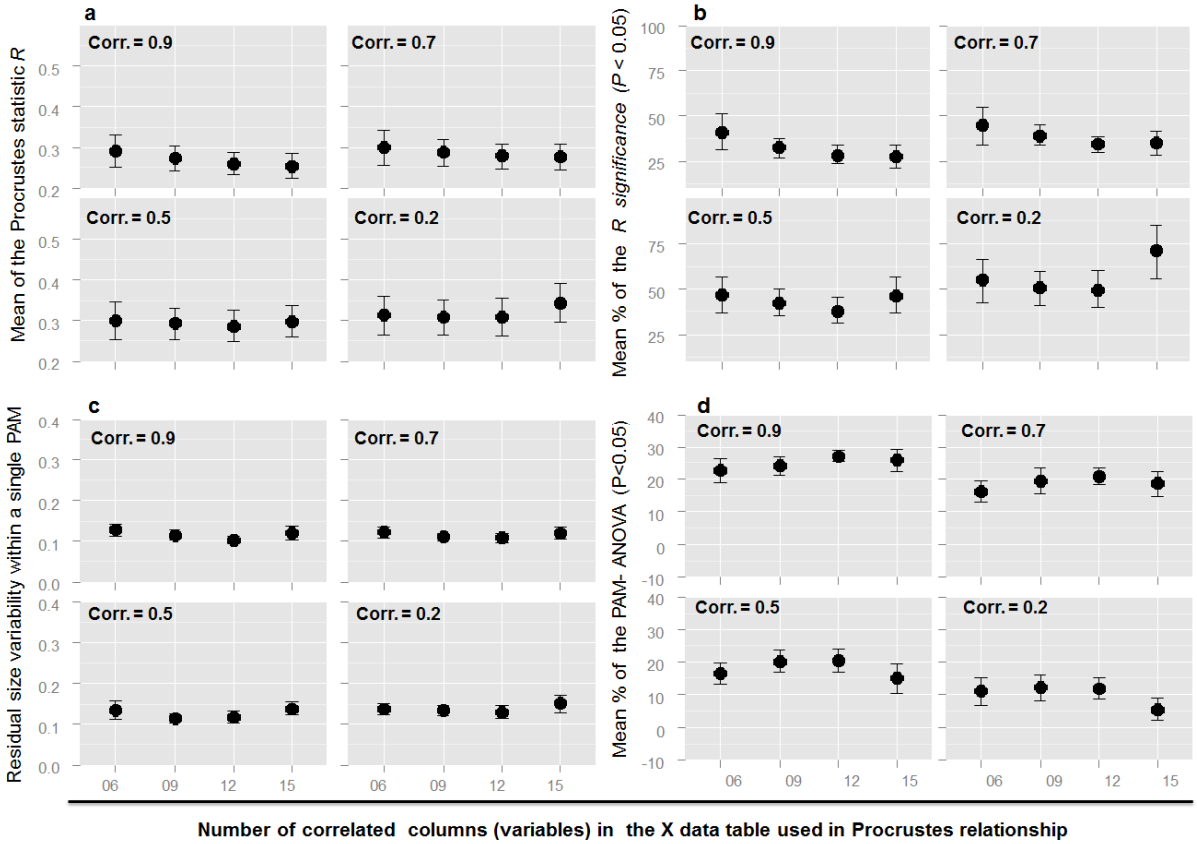


Figure 2. Effects of increasing number of correlated variables (6, 9, 12, 15) across in different levels (0.9; 0.7; 0.5; 0.2) within an entire **X** data table on Procrustes results.**a)** Effect on Procrustes correlation statistic *R*; **b)** Effect on significance of Procrustean

relationship (*P* value); **c)** Effect on residual size ranging within the vector of relationship (Procrustean association metric: PAM); **d)** Effect on ANOVA significance by using the PAMs as response and land use type (4 levels) as categorical predictor. The **X** (soil fertility) and the **Y** (PLFA profile) data tables are derived from Lisboa et al. (2014b). The correlation within **Y** data table was held fixed (original correlation structure). Means ± 1 SE of 13 pre-Procrustes transformations simulated 100 times are shown (Fig. 1c). Procrustes statistic *R* was negatively related to the number of significant PAM-ANOVA results(Fig. 5). The measure of residual size variability across individual PAMs, the residual ranging size within a single PAM, was negatively and consistently related to the number of significant PAM-ANOVA results, irrespective of the increasing number of correlated columns and the imposed correlation level (Fig. 5).

5.5.3. Correlation between mainstream and downstream results

Despite a trend of decreasing, the number of significant PAM-ANOVA results ($P < 0.05$) was not clear related to the increasing number of significant Procrustes statistic *R* ($P < 0.05$) across imposed correlation levels in **X** data table (Fig. 5). On other hand, within the highest correlation levels (0.9; 0.7; and 0.5), the higher Procrustes statistic *R* was related to the higher number of significant PAM-ANOVA results, irrespective of the number of correlated columns; however, in the lowest imposed correlated level (0.2) the increasing Procrustes statistic *R* was negatively related to the number of significant PAM-ANOVA results(Fig. 5). The measure of residual size variability across individual PAMs, the residual ranging size within a single PAM, was negatively and consistently related to the number of significant PAM-ANOVA results; irrespective of the increasing number of correlated columns and the imposed correlation level (Fig. 5).

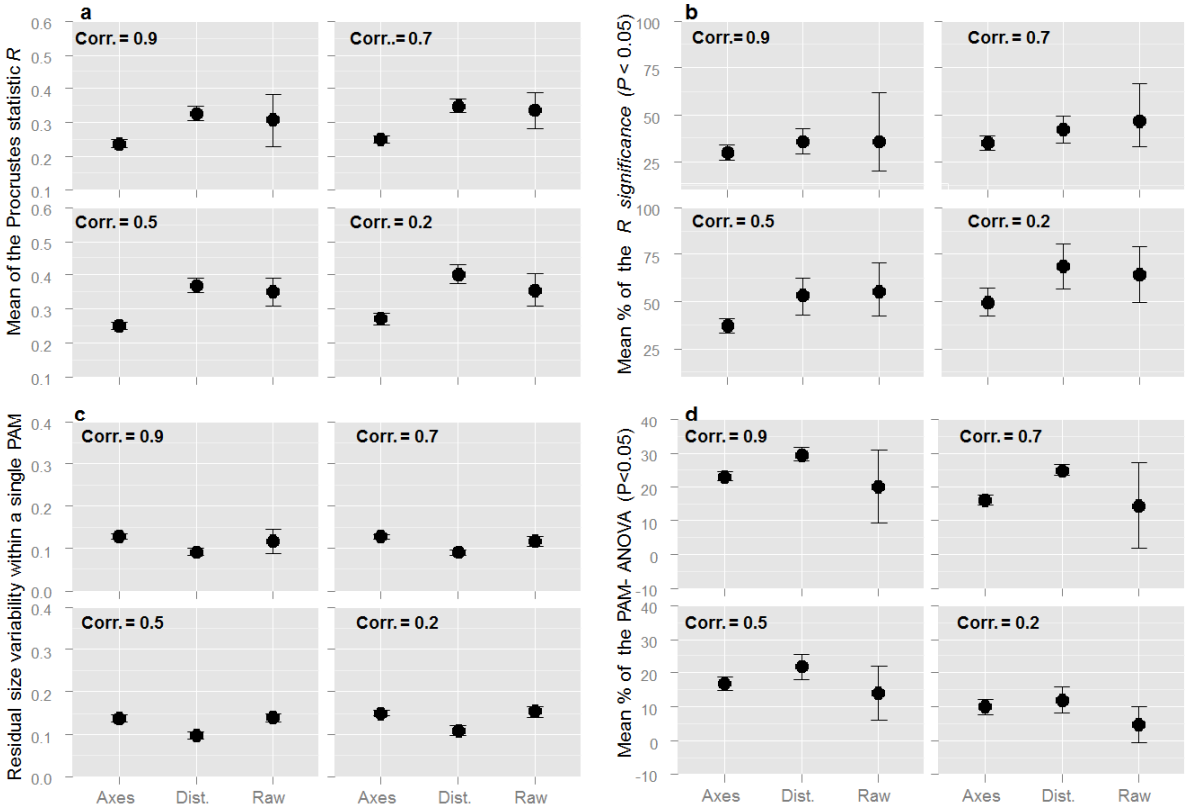


Figure 3. Effects of the increasing correlation level within a data table on Procrustes results by using different forms through which**X** and **Y** data table can be analyzed in Procrustes: raw data, distance matrices, and ordination axes. Each point represents different sets of the pre-transformations (Fig. 1c) corresponding to **X** and **Y** data tables in form of raw data, distance matrices, and ordination matrices, irrespective of the number of correlated columns.

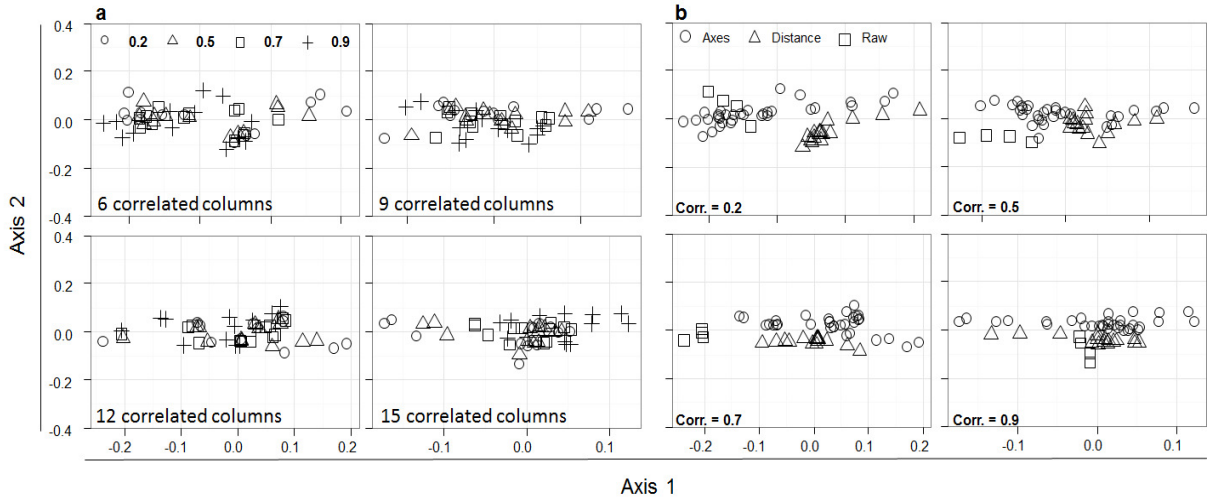


Figure 4.NMDS ordinations (Euclidian distance) of Procrustes residual vectors (PAMs). In **a**, PAMs are grouped according to the number of correlated variables in X data table. In **b**, the PAMs are grouped based on the form in wich **X** and **Y** data table were used in Procrustes (raw data, distance matrices, and ordination axes). Each symbol represents the mean PAM from 100 simulation of each pre-Procrustes treatment described in Fig. 1c.

5.5.4. Correlation between X and Y data tables for a specific treatment

The mean percentage of significant ANOVAs using PAMs as the response variable increased as the correlation level between **X** and **Y** data tables for treatment A increased (Fig. 6a). The mean number of times where the treatment A was significantly different from all other treatments (% A ≠ BCD) increased as the correlation level between **X** and **Y** for the treatment A increased (Fig. 6b). For all dimensions of **X** and **Y**, the higher correlation levels between these two data tables for the treatment A (0.7 and 0.9) were reflected by the mean Procrustes residual size for treatment A being lower than others B, C, and D (Fig. 7a-c). At the lower correlation levels between the **X** and **Y** data tables for the treatment A (0.2 and 0.5) the mean Procrustes residual size for treatment A was not different from the others, B, C, and D (Fig. 7a-c).

5.6. DISCUSSION

The use of the Procrustes residual vector (PAM) in ANOVA and multiple comparisons is not widespread as an ecological routine (Lisboa et al., 2014a). The reasons for this are diverse, including the lack of studies exploring the limitations of this composite framework. Here, we have attempted to address two questions: 1) how does the increasing number of columns correlation within an entire data table affects the Procrustes results? 2) Can the Procrustean residual vector (i.e. the PAM) come out differences among treatments in terms of multivariate correlation when it is used in multiple comparisons of means?

5.6.1. Correlation level within a data table does not clearly affect the Procrustes fit and PAM-ANOVA results

The most common use of Procrustes analysis in the ecological literature is for comparing different methodologies. For example, in soil microbiology Procrustes analysis has been used for comparing ordination patterns from different methods of accessing the soil microbial community (e.g. PLFA, T-RFLP, high throughput sequencing) (Vinten et al., 2011). Others authors have used Procrustes to assess how sampling error levels could affect the correlation between ordinations (Hirst and Jackson, 2007). All these examples used Procrustean parameters, such as the Procrustean *R* correlation statistic and its significance for assessing the fit between methodologies, which stress that the Procrustes fit is usually the final aim of the most part of studies using that approach.

The low appearance of Procrustes in the ecological literature, especially in plant and soil ecology (Lisboa et al., 2014a), results in no information on the consequences of the correlation within an entire data on Procrustes results and, in the sense of using the Procrustean residual vector (PAM) in others statistical frameworks, such as ANOVA, there are no any reference. Here our results indicated no clear effects of the increasing number of correlated variables across different imposed correlation levels in **X** data table.Procrustes

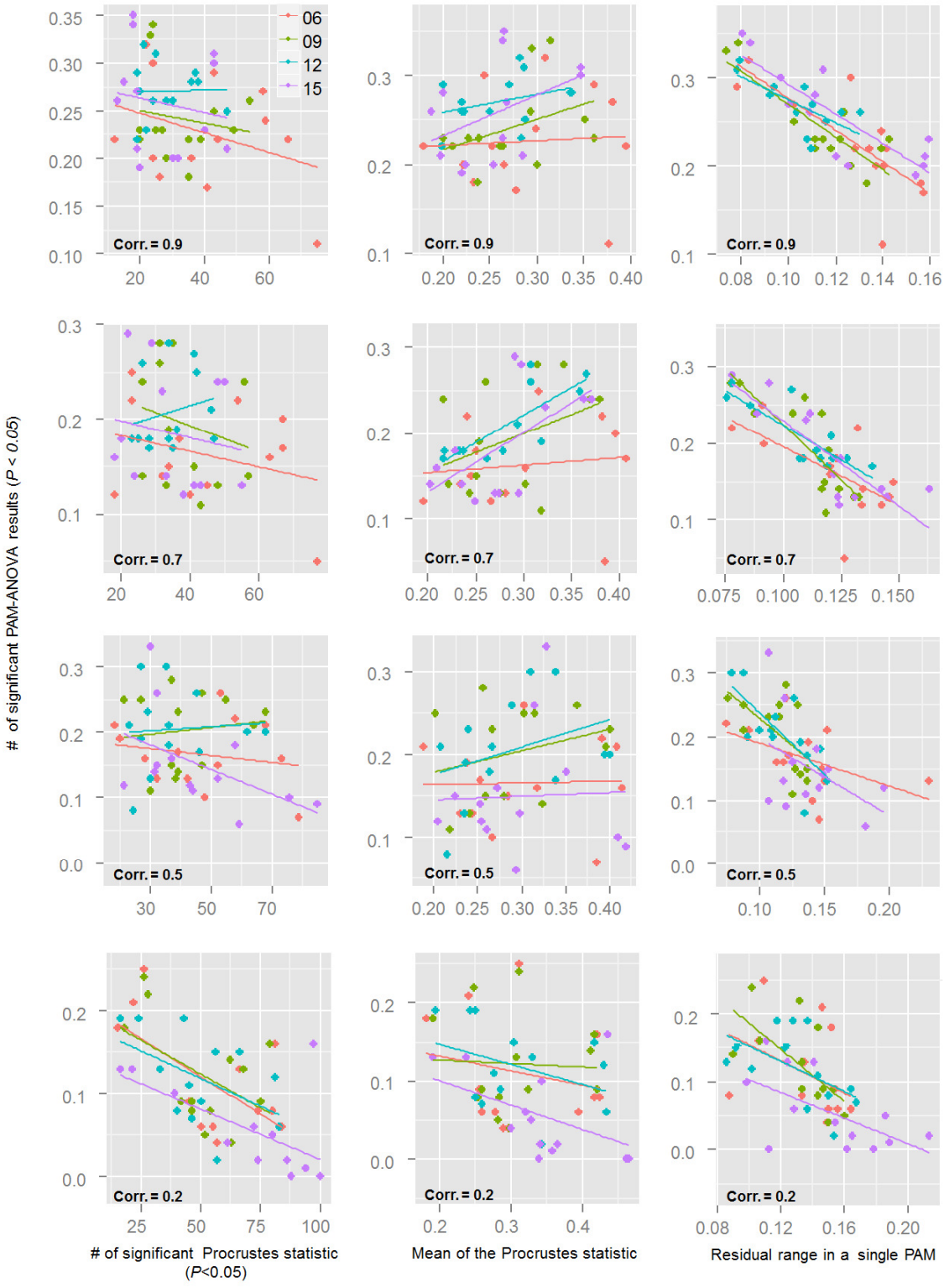


Figure 5. Correlation between Procrustes outcome related to the fit (mainstream Procrustes results) and the Procrustes outcome related to the use of the PAM in ANOVA framework (downstream result). The lines represent the numbers of correlated columns/variable within the X data table having a specific correlated level (0.9; 0.7; 0.5; and 0.2). Each point arises from a pre-transformation used for simulating the procrustes relationship between **X** and **Y** data table hundred times.

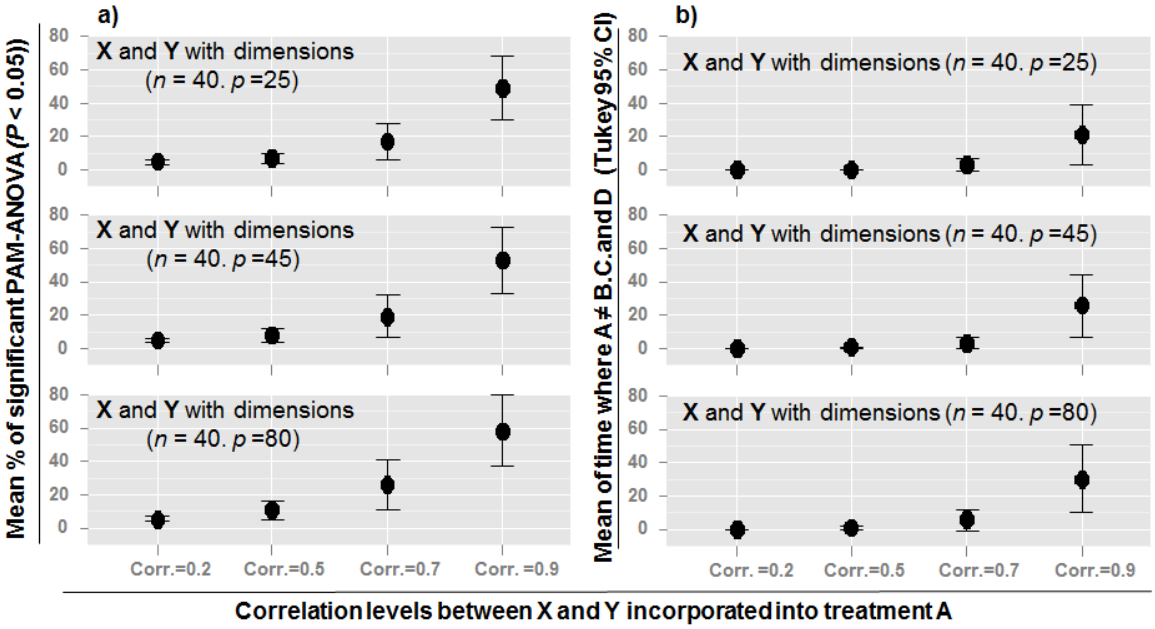


Figure 6. Procrustes downstream results as affected by the correlation levels between **X** and **Y** data tables incorporated into specific treatment A, while holding fixed the correlation level between **X** and **Y** for others treatments B, C, and D (correlation <0.1).**a**) Mean percentage of significant ANOVA ($P< 0.05$) when the Procrustes residual vectors (PAMs) were used as response variable. **b**) Mean percentage of the times that A treatment was significant different from all other treatments in multiple comparison (Tukey, 95%, CI). Means ± 1 SE of 13 pre-Procrustes transformations simulated 100 times are shown (Fig. 1c).

analysis used for carrying our simulation out was that occurring as default in vegan R package (Oksanen et al., 2013), which keeps one of the configurations fixed while the other configuration is submitted to translation and rotation in order to minimize the sum of squares of corresponding points. In our simulations the fixed configuration was that from **X** data table, that is, the data table with the increasing number of correlated columns across different imposed correlation levels. Thus all linear transformations which are inherent to Procrustes, such as translation to find a common centre between configurations followed by rotation on a constant angle to find the best fit between them, were only applied to **Y** data table, which in our simulation was that keeping its original correlation structure. In more simple words it means that the Procrustes "bed" was the correlation imposed data table, **X**,and the Procrustes"guest" was the non-imposed correlation data table, **Y**. Thereby, despite of both, **X** and **Y** data tables, have been submitted to the same set of Procrustes pre-transformations (Fig. 1c), only the **Y** data table – multidimensional configurations were object of the translation and rotation. It is important to be stressed as it has to do with the statement that Procrustes fit take only account the differences between **X** and **Y** configurations (Dray et al., 2003), which can be seen by:

$$fit_{XY} = \text{trace}\,((\mathbf{X} - \mathbf{Z})^T(\mathbf{X} - \mathbf{Z}))$$ (Gower, 1971),

where the **Z** matrix corresponds to the new set of coordinates arising from linear translation and rotation on **Y** data table – multidimensional configuration. It can cast light on our results showing that the increasing number of correlated columns across different imposed correlation levels did not come up with differences in terms of both, mainstream and downstream Procrustes results.

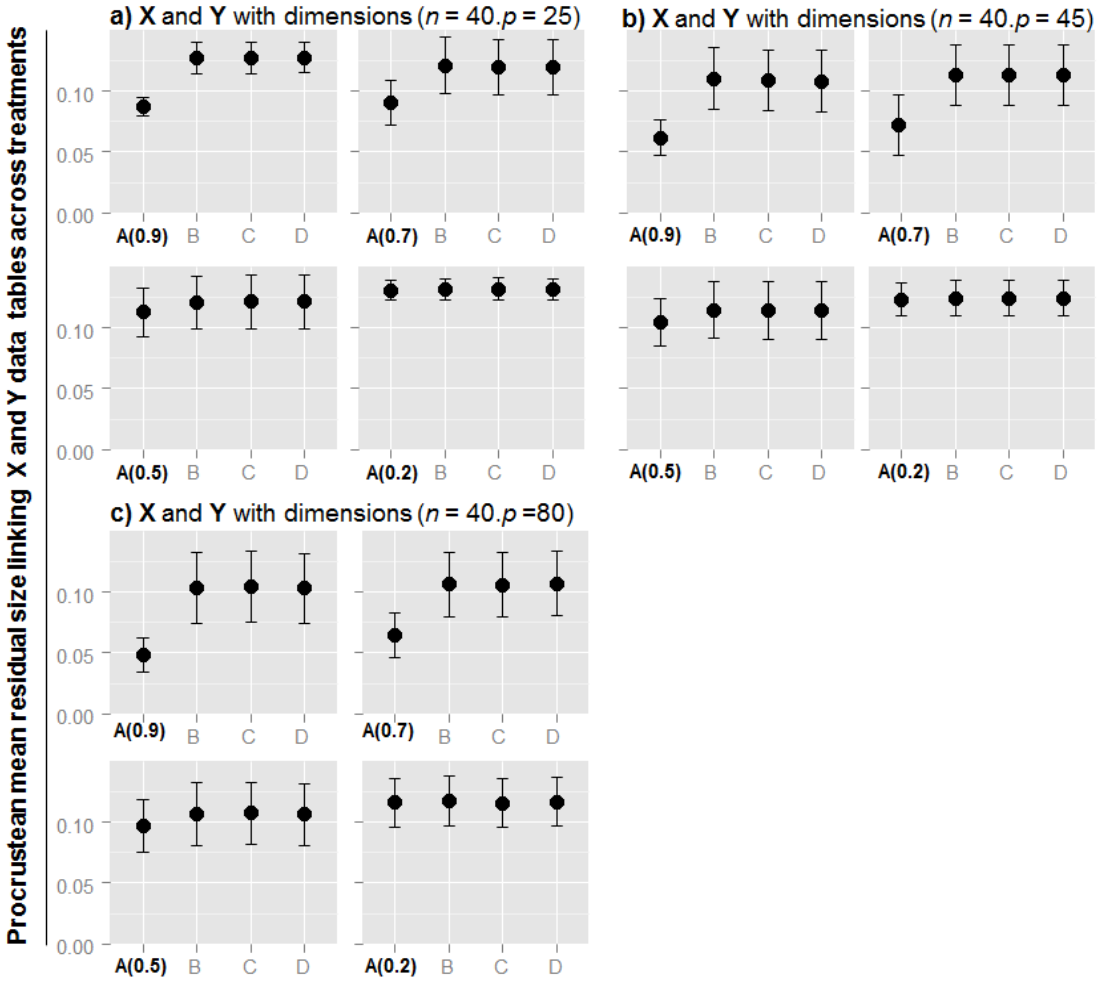


Figure 7. Accessing how multivariate correlation levels between two data tables incorporated into a specific treatment (treatment A: 0.9; 0.7; 0.5; and 0.2) are able to generated Procrustes residual vectors (PAMs) capable to differentiate this treatment from others (B, C, and D) in a multiple comparison. The correlation between X and Y for all other treatments (B, C, and D) was held fixed at <0.1.**a**) PAMs from Procrustes relationships between **X** and **Y** data tables with dimensions (n = 40, p =25), where n = nº rows and p = nº columns. **b)** PAMs from Procrustes relationships between X and Y data tables with dimensions (n = 40, p =45). **c:** PAMs from Procrustes relationships between X and Y data tables with dimensions (n = 40, p =80). All PAMs used in the multiple comparisons were generated from simulations of 13 pre-Procrustes analysis transformations as described in (Fig. 1c).

Since Procrustes takes "two to tango" by relating multidimensional configurations, which in turn are affected by the kind of pre-transformation on the data tables (Legendre and Gallagher, 2001), it would be expected that the **X** data table - multidimensional configurations (without translation and rotation) could have had some effect on Procrustes outcomes. Nonetheless, our results suggest there was no effect of the **X** data table - pre-transformations on their respective **X** data table – multidimensional configurations due to the high similarity between raw data, distance matrices and ordination axes in terms of Procrustes results and PAMs, irrespective of the imposed correlation level. Thus, this results suggests that the existing correlation within the non-translated and non-rotated configurations, in our case those from **X** data table, may not be a hurdle for the Procrustes results, irrespective of using raw, distance matrices, or ordination axes as entries.

5.6.2. Significant P value should not serve as a ground for choosing PAMs to be used in ANOVA

The use of the PAM in other statistical framework represents a non-well explored avenue of possibilities. It is which we can call of "analysis of the analysis" since we are analyzing the multivariate correlation depicted by a unique vector composed by Procrustean residuals, the PAM. One of the questions that could arise is whether the PAMs that should be uses on downstream statistical approaches such as ANOVA are only those coming from significant Procrustes relationship ($P < 0.05$). Does it really make sense? Our results showed that no. It is because the number of significant *R* value was not clear related to the number of significant PAM-ANOVA results across different imposed correlation levels (Fig. 5). On other hand, the other Procrustes fit outcome, the *R* statistic, was positively related to the number of significant PAM-ANOVA results, which suggest that, irrespective of significance, the higher correlation statistic the higher chances of a PAM-ANOVA significant result. However, it was observed that in the lowest imposed correlation level (0.2), the higher *R*, the lower is the chance of getting a significant PAM-ANOVA result, which indicates an existing correlation effect on the nature of the relationship between mainstream and downstream.

In addition, for another way of seeing the Procrustes fit: the residual size variation within a single PAM, we observed the clearest pattern of relationship between the numbers of significant PAM-ANOVA results (Fig. 5). As one can remember the residual sizes within a single PAM are connecting corresponding points across two multidimensional configurations after the best fit between them have been found out (Schneider and Borlund, 2007b). It means that the lower residual size variability, the higher is the fit between two configurations. Our results showed that independent from the imposed correlation level, and the number of columns correlating, the lower residual variability in a single PAM, the higher the number of significant PAM-ANOVA results (Fig. 5).Therefore, our results are showing that to use only PAMs coming from significant Procrustes relationship may not make sense.

5.6.3. What does PAM tell us?

An argument advocating the use of the Procrustes residual vector in a downstream statistical approach using ANOVA and multiple comparisons is that the consistencies in the Procrustean residual sizes, which are linking the two or more tables under investigation, could be used to make inferences on the strength of the multivariate correlation across environmental gradients. However, so far, no studies had explicitly explored such statement. In fact the few existing studies that used the PAM to make inference that goes beyond accessing the correlation between data tables were based on that statement (Singh et al., 2008, Landeiro et al., 2011, Siqueira et al. 2012, Lisboa et al., 2012, 2014b). Our results show that the correlation level between **X** and **Y** data tables for the treatment A affects the ANOVA and multiple comparisons of means using the PAM. We have found that the PAMs generated from the higher levels of correlation (0,7 and 0,9) are more capable of discriminating the treatment (A) from the others (B, C, and D). These results are supporting the statement in favor of using the PAM to assess the strength of the multivariate correlation across categorical levels.

One point that may raise confusion is the interpretation of the PAM-multiple comparisons of means results. We have used Tukey's HSD as it is a standard option in many studies, but we have found out that when only plotting the point estimate (mean), and the standard deviation of the residuals for each treatment, the pattern of lower residual size in highest correlation levels was clearer (Fig. 7). The assumptions are that since the link between two data tables is done through of residuals of the PAM after the best fit, then the lower residual, the higher is the multivariate correlation for a specific treatment. Our results support

this by showing that the mean residual size at treatment A was lower than the mean residual size in others treatments B, C, and D when the correlation between **X** and **Y** for treatment A was high (0,7 and 0,9). Thus, the overall interpretation for PAM using multiple comparisons is that low mean residual size for a treatment indicates "strong" multivariate correlation.

5.6.4. Final considerations

Here, we explored for the first time the effects of the increasing number of correlated columns across different imposed correlation levels within the X data table on the Procrustes results related to the fit (*R* statistic and its *P* value), and elated to the use of the Procrustean residual vector (PAM) in ANOVA. In addition, we also tried to show that the PAM when used in multiple comparisons can provide insights about differences among "treatments" in terms of multivariate relationship. We have only used data tables whose entries were quantitative, so only dissimilarities and transformations considered adequate for this kind of data were used in pre- transformation for getting the same dimension between **X** and **Y**. However, we do recognize that binary data tables (presence/absence) are also important in ecology (Anderson et al., 2011) and the evaluation of the correlation levels within binary data tables on Procrustes results must be an objective of future investigations.

Procrustes analysis is a symmetric approach used to link two or more data tables (Legendre and Legendre, 2012). This means that the data tables under analysis are evaluated on an equal footing, that is, without setting which of them is response or predictor. Also Procrustes does not have a regression step, which implies that the number of columns (variables, *p*) in a matrix does not need to be lower than the number of rows (sites, *n*) as required by traditional approaches to link data tables such as RDA (Redundancy analysis) and CCA (Canonical Correspondence Analysis). In the present study we did not do a formal evaluation to test the effects of $n>p$ on the Procrustes results as we focused on correlation effects. However, by varying the dimensions of **X** and **Y** data tables in the second part of the paper (n = 40, p = 25, 45, and 80) the results indicated that $n<p$ and $n>p$ may have similar effects on Procrustes results.

To our knowledge this is the first study exploring the correlation effects on the Procrustes results and interpretation. Here we showed that both the number of correlated variables and the correlation levels within an entire data table have no effects on in the mainstream Procrustes results related to the fit, such as *R* and its significance. In addition, the increasing correlating level within a data table does not affect the results of ANOVA using PAM as response. Overall, our study supports that the Procrustes fit only take into account the variation between data tables. However, we advocated that ecologists must be careful about the form in which the data tables to be investigated are used in the Procrustes analysis even our results have pointed no clear difference between data type entries (raw data, distance matrices, and ordination axes) in terms of Procrustes results. In addition, we found that using only PAMs from significant Procrustean relationships in downstream statistical analysis might not make sense since the results of PAM-ANOVA and multiple were not clearly correlated to the significance of the *R* statistics. It suggests that even PAMs from Procrustes analysis that came out as non-significant can be used in downstream analysis, such ANOVA and multiple comparisons of means, in order to evaluate the variation of the multivariate relationship across different categorical factors. Finally, we were able to show that the PAM do can reflect treatment differences in terms of multivariate correlation when it is used in mean multiple comparisons of means. It supports PAM – ANOVA – Multiple comparisons as an interesting composite approach for getting additional information on how the strength of the multivariate correlation varies across categorical environmental levels.

5.7. REFERENCES

ANDERSON, M. J.; CRIST, T. O.; CHASE, J. M.; VELLEND, M.; INOUYE, B. D.; FREESTONE, A. L.; SANDERS, N. J.; CORNELL, H, V.; COMITA, L. S.; DAVIES, K. F.; HARRISON, S. P.; KRAFT, N. J.; STEGEN, J. C.; SWENSON, N. G. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. **Ecology Letters**, v. 14, p. 19-28, 2011

DRAY, S.; CHESSEL, D.; THIOULOUSE, J. Procrustean Co-inertia analysis for the linking of multivariate datasets. **Ecoscience**,v. 10,p. 110-119, 2003

GOWER, J.C. Statistical methods of comparing different multivariate analyses on the same data. In: HODSON, F.R., KENDALL, D.G., TAUTU, P. editors. Mathematics in the archeological and historical sciences. Edinburgh University Press, Edinburgh. pp. 138-149, 1971

HARDIN, J.; GARCIA, S, R.; GOLAN, D.A method for generating realistic correlation matrices. **Annals of Applied Statistics**, v. 7, p. 1733-1762, 2013

HIRST, C. N.; JACKSON, D. A. Reconstructing community relationships: the impact of sampling error, ordination approach, and gradient length. **Diversity and Distribution**, v.13, p. 361-371, 2007

LANDEIRO, V. L.; BINI, L. M.; COSTA, F. R. C.; FRANKLIN, E.; NOGUEIRA, A.; DE SOUZA, L. P.; MORAES, J.; MAGNUSSON, W. E. How far can we go in simplifying biomonitoring assessments? An integrated analysis of taxonomy surrogacy, taxonomic sufficiency and numerical resolution in a mega diverse region. **Ecological Indicators**, v. 23, p. 366–373, 2012

LEGENDRE, P.; GALLAGHER, E. D. Ecologically meaningful transformations for ordination of species data.**Oecologia**, v. 129, p. 271–280, 2001

LEGENDRE, P.; LEGENDRE, L. Numerical Ecology, 3rd English edn.Elsevier344 Science BV, Amsterdam. 516 p. 483-495, 2012

LISBOA, F. J. G.; CHAER, G. M.; JESUS, E. C.; GONÇALVES, F. S.; SANTOS, F. M.; DE FARIA, S. M.; CASTILHO, A.; BERBARA, R. L. L. The influence of litter quality on the relationship between vegetation and below-ground compartments: a Procrustean approach. **Plant and Soil**, 367, p. 551-562, 2012

LISBOA, F. J. G.; PERES-NETO, P. R.; CHAER, G. M.; JESUS, E. C.; MITCHELL, R. J.; CHAPMAN, S. J.; BERBARA, R. L. L. Much beyond Mantel: Bringing Procrustes Association Metric to the Plant and Soil Ecologist's Toolbox. **PLoS ONE**, v. 9,p. e101238, 2014a doi:10.1371/journal.pone.0101238

LISBOA, F. J. G.; CHAER, G. M.; FERNANDES, M. F.; BERBARA, R. L. L.; MADARI, B. E. The match between microbial community structure and soil properties is modulated by land use types and sample origin within an integrated agroecosystem**. Soil Biology and Biochemistry**,v. 78: 97-108, 2014b

PERES-NETO, P. R.; JACKSON, D. A. How well multivariate data do sets match? The advantages of a Procrustean superimposition approach over the Mantel test. **Oecologia**,v. 129: 169-178, 2001

SCHNEIDER, J. W, BORLUND, P. Matrix Comparison, Part 2: measuring the resemblance between proximity measures or ordination results by use of the Mantel and Procrustes statistics. **Journal of the American Society for Information Science and Technology**v. 58, p.1596-1609, 2007

SINGH, B. K.; NUNAN, N.; RIDGWAY, K. P.; MCNICOL, J.; YOUNG, J. P. W.; DANIELL, T. J.; PROSSER, J. I.; MILLARD, P. Relationship between assemblages of mycorrhizal fungi and bacteria on grass roots. **Environmental Microbiology**, v. 10, p. 534–542, 2008

SIQUEIRA, T.; BINI, L. M.; ROQUE, F. O.; COTTIENE, K.A metacommunity framework for enhancing the effectiveness of biological monitoring strategies. **PLoS One**,v.7: e43626, 2012

VINTEN, A. J. I.; ARTZ, R. R. E.; THOMAS, N.; POTTS, J. M.; AVERY, L. M.; LANGAN, S. J.; WATSON, H.; COOK, Y.; TAYLOR, C.; ABEL, C. Comparison of microbial community assays for the assessment of stream biofilm ecology. **Journal of Microbiological Methods**, v. 85, p. 190-198, 2011

6. CONCLUSÕES FINAIS

Ao longo dos três capítulos da tese foram discutidos em detalhes o uso e a interpretação de resultados obtidos pelo uso da análise Procrustes. É preciso frisar, porém, que Procrustes é uma ferramenta para levantar indícios e não deve ser utilizada de forma exclusiva para fins de extrapolação dos resultados. Ou seja, Procrustes não é uma panacea. O mais prontamente visível inconveniente dessa abordagem é o fato de a correlação procrusteana ser expressa apenas em termos de magnitude e não em termos de natureza, isto é, Procrustes não informa se o relacionamento é negativo. Além disso, apesar do uso do vetor de resíduos, PAM, em outras abordagens estatísticas ter se mostrado como relativamente simples, o fato de essa abordagem ser aplicada em dois estádios pode inibir potenciais usuários, principalmente aqueles não versados em programas estatísticos que ofertam maior flexibilidade, como é o caso do programa R. Entretanto, esclarecidos tais inconvenientes, o conjunto da obra suportou a hipótese geral da tese de que a análise Procrustes pode ser utilizada como ferramenta útil para cientistas de solo e planta interessados em particionar a força de correlação multivariada em diferentes contextos de pesquisa. Ainda, potenciais usuários ganharam mais um suporte para diminuir suas ressalvas em relação à Procrustes uma vez que foi mostrado que os resultados da Procrustes são insensíveis ao aumento da correlação dentro das tabelas de dados sob análise. Somado a isso, a ideia de que o vetor de resíduos pode ser utilizado para verificar consistências dentro e entre coordenadas correspondentes foi sustentada pela corrente tese, mostrando que, de fato, a PAM é capaz de traduzir diferenças entre tratamentos em termos de correlação multivariada, aumentando a faixa de hipóteses potenciais que os cientistas de planta e solo podem testar em suas investigações.

Especificamente no estudo de caso do Capítulo II, que envolve dados do projeto CARBIOMA, a abordagem Procrusteana, ou Procrustes análises, associada à análise de variância e o teste de médias foi capaz de revelar que os sistemas de integração lavoura-pecuária-floresta divergem em relação aos outros tipos de uso da terra não apenas em temos de variáveis microbiológicas individuais, mas também em como os efeitos das propriedades do solo sobre a comunidade microbiana são modulados por ambos, tipo de uso da terra, e pela artificial heterogeneidade vegetal caracterizada pelo uso de espécies vegetais exóticas de forma sistematiza em meio às pastagens manejadas. Isso suportou a hipótese do caso de estudo sustentando que a relação fungo:bactéria, associada a ambientes de menor mineralização da matéria orgânica e, portanto, mais sustentável sob o ponto de vista de sequestro de carbono no solo, fosse mais destacada nos sistemas de integração do que nos outros tipos de uso da terra, em especial a pastagem degradada.

Contudo, é reconhecido que o presente estudo não foi de longo prazo, fator essencial para tornar mais realística a afirmativa de que os sistemas iLPF fomentam o domínio de fungos nos canais de transferência de energia entre solo e atmosfera. Dessa forma, com respeito ao estudo de caso do sistema de integração, muitas dúvidas permanecem uma vez que o foco foi apenas sobre uma das partes do binômio formado por estrutura e funcionamento do solo. Mesmo assim, o caso de estudo no Capítulo II foi o primeiro artigo a ter explicitamente como foco a variação em estrutura microbiana entre e dentro de sistemas de integração lavoura-pecuária floresta. Feitas as ressalvas, os resultados levantados pelo presente estudo de caso harmonizam com os postulados advogando a substituição das pastagens por sistemas de integração em busca de um novo paradigma de sustentabilidade para a agropecuária brasileira.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ACOSTA-MARTINEZ, V., BELL, C. W., MORRIS, B. E. L., ZAK, J., ALLEN, V. G. Long-term soil microbial community and enzymes activities response to an integrated cropping-livestock system in a semi-arid region. **Agriculture, Ecosystems and Environment** v. 137, p. 231-240, 2010.

BÅÅTH, E. The use of Neutral Lipids Fatty Acid to Indicate the Physiological Conditions of Soil Fungi. **Microbial Ecology**, v. 45, p. 373-383, 2003.

BALBINO, L.; CORDEIRO, L. A. M.; OLIVEIRA, P.; KLUTHCOUSKI, J.; GALERANI, P. R.; VILELA, L. Agricultura sustentável por meio da integração lavoura-pecuária-floresta (iLPF). **Informações agronômicas**, v. 138, p. 1-18, 2012

BALDRIAN, P. Fungal laccaes-occurrence and properties.**FEMS Microbiology Review**, v. 30, p. 215-242, 2006.

BARDGETT, R. D., JONES, A. C., JONES, D. L., KEMMITT, S. J., COOK, R., HOBBS, P. J. Soil microbial community patterns related to the history and intensity of grazing in sub-montane ecosystems. **Soil Biology and Biochemistry** v. 33, p. 1653-1664, 2001.

BARDGETT, R. D; MCALLISTER, E. The measurement of soil fungal: bacterial biomass ratios as an indicator of ecosystem self-regulation in temperate meadow grasslands. **Biology and Fertility Soils**, v.29, p. 282-290, 1999.

BODELIER, P. L. E.; GILLISEN, M. –J. B.; HORDIJK, K.; SINNINGHE, D.; RIJPSTRA, W. I. C.; GEENEVASEN, J. A. L.; DUNFIELS, P. F. A reanalysis of phospholipids fatty acid as ecological biomarkers for methanotrophic bacteria.**ISME Journal**, v. 3, p. 606-617, 2009.

BOSSIO, D. A.; SCOW, K. M..; GUNAPALA, N.; GRAHAM, K. J. Determinants of Soil Microbial Communities: Effects of Agricultural Management, Season, and Soil Type on Phospholipid Fatty Acid Profiles. **Microbial Ecology**, v. 36, p. 1–12, 1998

CHEN, J.; FERRIS, H.; SCOW, K. M.; GRAHAM, K. J. Fatty acid composition and dynamics of selected fungal-feeding nematodes and fungi. **Comparative Biochemistry and Physiology**, v. 130, p. 135-144, 2001.

CONNANT, R.; PAUSTIAN, K.; ELLIOT, E. T. Grassland management and conversion into grassland: effects on soil carbon. **Ecology**, v. 11, p. 343-355, 2001

DE DEYN, G. B.; CORNELISSEN, J. H. C. BARDGETT, R. D. Plant functional traits and soil carbon sequestration in contrasting biomes. **Ecology Letters**, v. 11, p. 516–531, 2008

DE VRIES, F. T.; BARDGETT, R. D. Plant–microbial linkages and ecosystem nitrogen retention: lessons for sustainable agriculture. **Frontiers in Ecology and Environment**, v.10, p. 425–432, 2012.doi:10.1890/110162

DE VRIES, F. T.; VAN GROENIGEN J. W.; HOFFLAND, E., BLOEM, J. Nitrogen losses from two grassland soils with different fungal biomass. **Soil Biology and Biochemistry**, v. 43, p. 997-1005, 2011.

DIAS-FILHO, M. B.; ANDRADE, C. M. S. Pastagens no trópico úmido. **Documentos 241**. Embrapa Amazônia Oriental, 2006

DOWLING, N. J. E.; WIDDEL, F.; WHITE, D. C. Phospholipid ester-linked fatty acid biomarkers of acetate-oxidizing sulphate-reducers and other sulphide-forming bacteria. **Journal of General Microbiology**, v. 132, p. 1815-1825, 1986.

FERNANDES, M. F.; BARRETO, A. C.; MENDES, I. C.; DICK, R. P. Short-term response of physical and chemical aspects of soil quality of a kaolinitic Kandiudalfs to agricultural practices and its association with microbiological variables. **Agriculture, Ecosystems & Environment**, v. 142, p. 419-427, 2011

FROSTEGÅRD, A.; BÅÅTH, E. The use of phospholipid fatty acid analysis to estimate bacterial and fungal biomass in soil. **Biology Fertility Soils**, v. 22, p. 59-65, 2001.
FROSTEGÅRD, Å.; TUNLID, A.; BÅÅTH, E. Use and misuse of PLFA measurements in soils. **Soil Biology & Biochemistry**, v. 43, p. 1621-1625, 2011

GOWER, J C. Statistical methods of comparing different multivariate analyses on the same data. In: Hodson FR, Kendall DG, Tautu P, editors. Mathematics in the archeological and historical sciences. Edinburgh University Press, Edinburgh. pp. 138-149

GUCKERT, J. B.; RINGELBERG, D. B.; WHITE, D. C.; HANSON, R. S.; BRATINA, B. J. Membrane fatty acids as phenotypic markers in the polyphasic taxonomy of methylotrophs within the Proteobacteria. **Journal of General Microbiology**, v. 137, p. 2631-2641, 1991.

GUO, L. B.; GIFFORD, R. M. Soil carbons stocks and land use changes: a meta-analysis. **Global Change Biology**, v. 8, p. 345-360, 2002.

HOSSAIN, Z.; SUGIYAMA, S- I. Geographical structure of soil microbial communities in northern Japan: Effects of distance, land use type and soil properties. **European Journal of Soil Biology**, v. 47, p. 88-94, 2011

KERGER, B. D.; NICHOLS, P. D.; ANTWORTH, C. P.; SAND, W.; BOCK, E.; COX, J. C; LANGWORTHY, T. A.; WHITE, D. C. Signature fatty acids in the polar lipids of acid-producing *Thiobacillus* spp.: methoxy, cyclopropyl, alpha-hydroxy-cyclopropyl, branched and normal monoenoic fatty acids. **FEMS Microbiology Ecology**, v. 38, p. 67-77, 1986.

LAL, R., FOLLETT, F., STEWART, B.A., KIMBLE, J.M. Soil carbon sequestration to mitigate climate change and advance food security. **Soil Science** v. 172, p. 943-956, 2007.

LANGER, U.; RINKLEBE, J. Priming effect after glucose amendment in two different soil evaluated by SIR-and PLFA-technique. **Ecological Engineering**, v. 37, p. 465-473, 2011.

LAPOLA, D. M.; MARTINELLI, L. A.; PERES, C. A.; OMETTO, J. P. H. B.; FERREIRA, M. E.; NOBRE, C. E.; D.AGUIAR, A. P.; BUSTAMANTE, M. C.; CARDOSO, M. F.; COSTA, M. H.; JOLY, C. A.; LEITE, C. C.; MOUTINHO, P.; SAMPAIO, G.; STRASSBURG, B. B. N.; VIEIRA, I. C. G. Pervasive transition of the Brazilian land-use system. **Nature Climate Changes**, v. 24, p. 27-35, 2014

LEE, A. K. Y.; CHAN, C. K.; FANG, M.; LAU, A. P. S. The 3-hydroxy fatty acids as biomarkers for quantification and characterization of endotoxins and Gram-negative bacteria

in atmospheric aerosols in Hong Kong. **Atmospheric Environment**, p. 38, p. 6307-6317, 2004.

LUNDIQUIST, E. J.; SCOW, K. M.; JACKSON, L. E.; UESUGI, S. L.; JOHNSON, C. R. Rapid response of soil microbial communities from conventional, low input, and organic farming systems to a wet/dry cycle. **Soil Biology & Biochemistry**, v. 31, p. 1661-1675, 1999

MADAN, R.; PANKHURST, C.; HAWKE, B.; SMITH, S. Use of fatty acid for identification of AM fungi and estimation of the biomass of spores in soil. **Soil Biology & Biochemistry**, v. 34, p. 125-128, 2002.

MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, v. 27, p. 209-220, 1967

MILLARD, P., SINGH, B. K. Does grassland drive soil microbial diversity? **Nutrient Cycling in Agroecosystems**, v. 88, p. 147-158, 2010.

MIRZA, M. S.; JANSE, J. D.; HAHNN, D.; AKKERMANS, A. D. L. Identification of atypical Frankia strains by fatty acid analysis. **FEMS Microbiology Letters**, v. 83, p. 91-98, 1991

MOESKOPS, B., BUCHAN, D., SUKRISTIYONUBOWO., DE NEVE, S., DE GUSSEME, B., WIDOWATI, L. R., SETYORINI, D., SLEUTEL S. Soil quality indicators for intensive vegetable procuctions systems in Java, Indonesia. **Ecological Indicators**, v. 18, p. 218-226, 2012.

NEIDHARDT, F. C.; INGRAHAM, J. L.; SCHAECHTER, M. Physiology of the Bacterial Cell: A Molecular Approach. Sinauer Associates, Sunderland.MA, 1990.

OLSSON, P. A. Signature fatty acids provide tools for determination of the distribution and interactions of mycorrhizal fungi in soil. **FEMS MicrobiologyEcology**, v. 29, p. 303-310, 1999.

PARTON, W. J., SCURLOCK, J. M. O., OJIMA, D. S., GILMANOV, T.G., SCHOLES, R.J., SCHIMEL, D.S., KIRCHNER, T., MENAUT, J.C., SEASTEDT, T., MOYA, E.G., KAMNALRUT, A., KINYAMARIO, J.I. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. **Global Biogeochemical Cycles,** v. 7, p. 785-809, 1993.

PINKART, H.C.; RINGELBERG, D.B.; PICENO, Y.M.; MACNAUGHTON, S.J.; WHITE, D.C. Biochemical approaches to biomass measurements and community structure analysis, in: HURST, C.J.; CRAWFORD, R.L.; KNUDSEN, G.R.; MCINERNEY, M.J.; STETZENBACH, L.D. (Eds.), **Manual of Environmental Microbiology**. Washington DC: American Society for Microbiology Press, p. 101–113. 2002.

POTTHAST, H.; HAMMER, U.; MAKESCHIN, F. In an Ecuadorian pasture soil the growth of *Setaria sphacelata*, but not of soil microorganisms, is co-limited by N and P. **Applied Soil Ecology**, v. 62, p. 103-114, 2012

POTTHAST, K., HAMER, U., MAKESCHIN, F. Land use change in a tropical mountain rainforest region of southern Ecuador affects soil microorganisms and nutrient cycling. **Biogeochemistry**, v. x, p. 1–17, 2011

RAMSEY, P. W.; GIBBONS, S. M.; RICE, P.; MUMMEY, D.; FERIS, K. P.; MOORE, J. N.; RILLIG, M. C.; GANNON, J. E. Relative strengths of relationships between plant, microbial, and environmental parameters in heavy-metal contaminated floodplain soil. **Pedobiologia**, v. 55, p. 15-23

RINGELBERG, D. B.; DAVIS, J. D.; SMITH, G. A.; PFIFFNER, S. M.; NICHOLS, P. D.; NICKELS, J. S.; HENSON, J. M.; WILSON, J. T.; YATES, M.; KAMPBELL, D. H.; READ, H. W.; STOCKSDALE, T. T.; WHITE, D. C. Validation of signature polar lipid fatty acid biomarkers for alkane-utilizing bacteria in soils and subsurface aquifer materials. **FEMS Microbiology Ecology**, v. 62, p. 39-50, 1989.

RINNAN, R.; BÅÅTH, E. Differential utilization of carbon substrates by bacteria and fungi in tundra soil. **Applied and Environmental Microbiology**, v. 75, p. 3611-3620, 2009.

RODRIGUES, J. L.; PELLIZARI, V.H.; MUELLER, R.; BAEK, K.; JESUS, EDA, C.; PAULA, F. S.; MIRZA, B.; HAMAOUI, G. S. JR.; TSAI, S. M.; FEIGL, B.; TIEDJE, J. M.; BOHANNAN, B. J.; NUSSLEIN, K. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proceedings of the National Academy of Sciences**, v. 110, p. 988-993, 2013

ROUSK, J., BAATH, E. Fungal and bacterial growth in soil with plant materials of different C/N ratios. **FEMS Microbiology Ecology**, v. 62, p. 258-267, 2007.

RUESS, L.; CHAMBERLAIN, P. M. The fat that matters: Soil food web analysis using fatty acid and their carbon stable isotope signature. **Soil Biology & Biochemistry**, v. 42, p. 1898-1910, 2010.

RUESS, L.; GARCÍA ZAPATA, E. J.; DIGHTON, J. Food preferences of a fungal-feeding Aphelenchoides species. **Nematology**, v. 2, p. 223-230, 2000.

RUESS, L.; SCHÜTZ, K.; MIGGE-KLEIAN, S.; HÄGGBLOM, M. M.; KANDELER, E.; SCHEU, S. Lipid composition of Collembola and their food resources in deciduous forest stands - implications for feeding strategies. **Soil Biology & Biochemistry**, v. 39, p. 1990-2000, 2007.

SAKAMOTO, K.; LIJIMA, T.; HIGUCHI, R. Use of specific phospholipid fatty acid for identifying and quantifying the external hyphae of the arbuscular mycorrhizal fungus *Gigaspora rosea*. **Soil Biology & Biochemistry**, v. 36, p. 1827-1834, 2004.

SCHARROBA, A.; DIBBERN, D.; HUNNINGHAUS, M.; KRAMER, S.; MOLL, J.; BUTENSCHOEN, O.; BONKOWSKI, M.; BUSCOT.; KANDELER, E.; KOLLER, KRUGGER, D.; LUEDERS, T.; SCHEU, S.; RUESS, L. Effects of resource availability and quality on the structure of the micro-food web of an arable soil across depth. **Soil Biology & Biochemistry**, v. 50, p. 1-11, 2012

SCHLESINGER, W. Soil organic matter: a source of atmospheric CO_2. In: WOODWELL, G. M. (ed) The role of terrestrial vegetation in global carbon cycle, 1994

SIX, J.; FREY, S. D.; THIET, R. K.; BATTEN, K. M. Bacterial and Fungal Contributions to Carbon Sequestration in Agroecosystems. **Soil Science Society American. Journal**, v. 70, p. 555–569, 2006

STRASSBURG, B. B. N.; LATAWIEC, A.; BARIONE, L. G.; NOBRE, C. A.; DA SILVA, W.; VALENTIM, J. F.; ASSAD, E. D. VIANNA, M.. When enough should be enough: Improving the use of current agricultural lands could meet production demands and spare natural habitats in Brazil, **Global Environment Change**, v. 28, p. 84-97, 2014

VALLEJO, V.E., ARBELI, Z., TERÁN, W., LORENZ, N., DICK, R.P., ROLDAN, F., 2012. Effect of land management and *Prosopis juliflora* (Sw.) DC trees on soil microbial community and enzymatic activities in intensive silvopastoral systems of Colombia. **Agriculture, Ecosystems & Environment**, v. 150, p. 139-148.

VAN DER HEIJDEN, M. G. A.; BARDGETT, R. D.; VAN STRAALEN, N. M. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. **Ecology Letters**, v. 11, p. 296-310, 2008.

VAN GROENIGEM, K. J.; BLOEM, J.; BAATH, E.; BOECKX, P.; ROUSK, J.; BODÉ, S.; FORRISTAL, D.; JONES, M. B. Abundance, production and stabilization of microbial biomass under conventional and reduced tillage. **Soil Biology & Biochemistry**, v. 42, p. 48-55, 2010.

WAKEHAM, S. G.; PEASE, T. K.; BENNER, R. Hydroxy fatty acids in marine dissolved organic matter as indicators of bacterial membrane material. **Organic Geochemistry**, v. 34, p. 857-868, 2003.

WALLENIUS, K.; RITA, H.; MIKKONEN, A.; LAPPI, K.; LINDSTROM, K.; HARTKAINEN, H.; RAATLAND, A.; NIEMI, R. M. Effects of land use on the level, variation and spatial structure of soil enzyme activities and bacterial communities. **Soil Biology & Biochemistry**, v. 43, p. 1464-1473, 2011

WESTHUIZEN VAN DER, J. P. J.; KOCK, J. L. F.; BOTHA, A.; BOTES, P. J. The distribution of the u3 and u6 series of cellular long-chain fatty acids in fungi. **Systematic Applied Microbiology**, v. 17, p. 327-345, 1994.

ZAADY, E.; BEN-DAVID, E. A.; SHER, Y.; TZIRKIN, R.; NEJIDAT, A. Inferring biological soil crust sucessional stage using combined PLFA, DGGE, physical and biophysiological analyses. **Soil Biology & Biochemistry**, v. 42, p. 842-849, 2010.

ZELLES, L. Fatty acid patterns of phospholipids and lipopolysaccharides in the characterization of microbial communities in soil: a review. **Biology Fertility and Soils**, v.29, p. 111-129, 1999.

ZELLES, L. Phospholipid fatty acid profiles in selected members of soil microbial communities. **Chemosphere**, v. 35, p. 275-294, 1997.