

# Aplicações da Linguagem R em Análises de Vegetação

---

Écio Souza Diniz  
Pedro Manuel Villa  
(Organizadores)



# Aplicações da Linguagem R em Análises de Vegetação

---

Écio Souza Diniz  
Pedro Manuel Villa  
(Organizadores)

2020 by Atena Editora

Copyright © Atena Editora

Copyright do Texto © 2020 Os autores

Copyright da Edição © 2020 Atena Editora

**Editora Chefe:** Profª Drª Antonella Carvalho de Oliveira

**Diagramação:** Geraldo Alves

**Edição de Arte:** Lorena Prestes

**Revisão:** Os Autores



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição *Creative Commons*. Atribuição 4.0 Internacional (CC BY 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

### **Conselho Editorial**

#### **Ciências Humanas e Sociais Aplicadas**

Profª Drª Adriana Demite Stephani – Universidade Federal do Tocantins

Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas

Prof. Dr. Alexandre Jose Schumacher – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso

Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná

Prof. Dr. Antonio Gasparetto Júnior – Instituto Federal do Sudeste de Minas Gerais

Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília

Prof. Dr. Carlos Antonio de Souza Moraes – Universidade Federal Fluminense

Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa

Profª Drª Cristina Gaio – Universidade de Lisboa

Profª Drª Denise Rocha – Universidade Federal do Ceará

Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia

Prof. Dr. Edvaldo Antunes de Farias – Universidade Estácio de Sá

Prof. Dr. Eloi Martins Senhora – Universidade Federal de Roraima

Prof. Dr. Fabiano Tadeu Grazioli – Universidade Regional Integrada do Alto Uruguai e das Missões

Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná

Profª Drª Ivone Goulart Lopes – Istituto Internazionale delle Figlie di Maria Ausiliatrice

Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense

Profª Drª Keyla Christina Almeida Portela – Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso

Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins

Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte

Prof. Dr. Marcelo Pereira da Silva – Universidade Federal do Maranhão

Profª Drª Miranilde Oliveira Neves – Instituto de Educação, Ciência e Tecnologia do Pará

Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa

Profª Drª Rita de Cássia da Silva Oliveira – Universidade Estadual de Ponta Grossa

Profª Drª Sandra Regina Gardacho Pietrobon – Universidade Estadual do Centro-Oeste

Profª Drª Sheila Marta Carregosa Rocha – Universidade do Estado da Bahia

Prof. Dr. Rui Maia Diamantino – Universidade Salvador

Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará

Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

Prof. Dr. William Cleber Domingues Silva – Universidade Federal Rural do Rio de Janeiro

Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

#### **Ciências Agrárias e Multidisciplinar**

Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano

Prof. Dr. Antonio Pasqualetto – Pontifícia Universidade Católica de Goiás

Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná

Profª Drª Diocléa Almeida Seabra Silva – Universidade Federal Rural da Amazônia  
Prof. Dr. Écio Souza Diniz – Universidade Federal de Viçosa  
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul  
Prof. Dr. Fágner Cavalcante Patrocínio dos Santos – Universidade Federal do Ceará  
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia  
Prof. Dr. Júlio César Ribeiro – Universidade Federal Rural do Rio de Janeiro  
Profª Drª Lina Raquel Santos Araújo – Universidade Estadual do Ceará  
Prof. Dr. Pedro Manuel Villa – Universidade Federal de Viçosa  
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão  
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará  
Profª Drª Talita de Santos Matos – Universidade Federal Rural do Rio de Janeiro  
Prof. Dr. Tiago da Silva Teófilo – Universidade Federal Rural do Semi-Árido  
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

### **Ciências Biológicas e da Saúde**

Prof. Dr. André Ribeiro da Silva – Universidade de Brasília  
Profª Drª Anelise Levay Murari – Universidade Federal de Pelotas  
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás  
Prof. Dr. Edson da Silva – Universidade Federal dos Vales do Jequitinhonha e Mucuri  
Profª Drª Eleuza Rodrigues Machado – Faculdade Anhanguera de Brasília  
Profª Drª Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina  
Prof. Dr. Ferlando Lima Santos – Universidade Federal do Recôncavo da Bahia  
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria  
Prof. Dr. Igor Luiz Vieira de Lima Santos – Universidade Federal de Campina Grande  
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará  
Profª Drª Magnólia de Araújo Campos – Universidade Federal de Campina Grande  
Profª Drª Mylena Andréa Oliveira Torres – Universidade Ceuma  
Profª Drª Natiéli Piovesan – Instituto Federaci do Rio Grande do Norte  
Prof. Dr. Paulo Inada – Universidade Estadual de Maringá  
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa  
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande

### **Ciências Exatas e da Terra e Engenharias**

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto  
Prof. Dr. Alexandre Leite dos Santos Silva – Universidade Federal do Piauí  
Prof. Dr. Carlos Eduardo Sanches de Andrade – Universidade Federal de Goiás  
Profª Drª Carmen Lúcia Voigt – Universidade Norte do Paraná  
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná  
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará  
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande  
Prof. Dr. Marcelo Marques – Universidade Estadual de Maringá  
Profª Drª Neiva Maria de Almeida – Universidade Federal da Paraíba  
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte  
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

### **Conselho Técnico Científico**

Prof. Msc. Abrãao Carvalho Nogueira – Universidade Federal do Espírito Santo  
Prof. Msc. Adalberto Zorzo – Centro Estadual de Educação Tecnológica Paula Souza  
Prof. Dr. Adailson Wagner Sousa de Vasconcelos – Ordem dos Advogados do Brasil/Seccional Paraíba  
Prof. Msc. André Flávio Gonçalves Silva – Universidade Federal do Maranhão  
Profª Drª Andreza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico  
Profª Msc. Bianca Camargo Martins – UniCesumar  
Prof. Msc. Carlos Antônio dos Santos – Universidade Federal Rural do Rio de Janeiro  
Prof. Msc. Cláudia de Araújo Marques – Faculdade de Música do Espírito Santo  
Prof. Msc. Daniel da Silva Miranda – Universidade Federal do Pará  
Profª Msc. Dayane de Melo Barros – Universidade Federal de Pernambuco



Prof. Dr. Edwaldo Costa – Marinha do Brasil  
 Prof. Msc. Eliel Constantino da Silva – Universidade Estadual Paulista Júlio de Mesquita  
 Prof. Msc. Gevair Campos – Instituto Mineiro de Agropecuária  
 Prof. Msc. Guilherme Renato Gomes – Universidade Norte do Paraná  
 Prof<sup>a</sup> Msc. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia  
 Prof. Msc. José Messias Ribeiro Júnior – Instituto Federal de Educação Tecnológica de Pernambuco  
 Prof. Msc. Leonardo Tullio – Universidade Estadual de Ponta Grossa  
 Prof<sup>a</sup> Msc. Lilian Coelho de Freitas – Instituto Federal do Pará  
 Prof<sup>a</sup> Msc. Liliani Aparecida Sereno Fontes de Medeiros – Consórcio CEDERJ  
 Prof<sup>a</sup> Dr<sup>a</sup> Lívia do Carmo Silva – Universidade Federal de Goiás  
 Prof. Msc. Luis Henrique Almeida Castro – Universidade Federal da Grande Dourados  
 Prof. Msc. Luan Vinicius Bernardelli – Universidade Estadual de Maringá  
 Prof. Msc. Rafael Henrique Silva – Hospital Universitário da Universidade Federal da Grande Dourados  
 Prof<sup>a</sup> Msc. Renata Luciane Polsaque Young Blood – UniSecal  
 Prof<sup>a</sup> Msc. Solange Aparecida de Souza Monteiro – Instituto Federal de São Paulo  
 Prof. Dr. Welleson Feitosa Gazel – Universidade Paulista

**Dados Internacionais de Catalogação na Publicação (CIP)  
(eDOC BRASIL, Belo Horizonte/MG)**

A642 Aplicações da linguagem R em análises de vegetação [recurso eletrônico] / Organizadores Écio Souza Diniz, Pedro Manuel Villa. – Ponta Grossa, PR: Atena, 2020.

Formato: PDF  
 Requisitos de sistema: Adobe Acrobat Reader  
 Modo de acesso: World Wide Web  
 Inclui bibliografia  
 ISBN 978-65-86002-35-5  
 DOI 10.22533/at.ed.355200903

1. Desenvolvimento sustentável. 2. R (Linguagem de programação de computador). 3. Recursos vegetais – Manejo.  
 I. Diniz, Écio Souza. II. Villa, Pedro Manuel.

CDD 333.7511

**Elaborado por Maurício Amormino Júnior – CRB6/2422**

Atena Editora  
 Ponta Grossa – Paraná - Brasil  
[www.atenaeditora.com.br](http://www.atenaeditora.com.br)  
 contato@atenaeditora.com.br

## APRESENTAÇÃO

Os diferentes tipos de vegetação ao redor do globo, principalmente as florestas tropicais, se destacam por fornecer importantes bens e serviços ecossistêmicos para a humanidade como, por exemplo, regulação climática, provisão de alimentos e diversas fontes de energia. Contudo, as crescentes e rápidas mudanças no meio ambiente causadas por sua intensa exploração no século 21 têm promovido reduções drásticas de importantes vegetações distribuídas em distintos Biomas. O Brasil como um país de dimensão continental e rico em recursos vem atravessando profundas transformações em seus Biomas, o que é destacadamente devido aos usos intensos da terra sem técnicas adequadas de manejo para a sua exploração.

Diante desse panorama de significativas transformações do meio natural, se faz necessário e urgente o estudo de diferentes tipos de comunidades vegetais e ecossistêmicas para definir estratégias de manejo e conservação, assim como pesquisas que visem a otimização de produções agrícolas de forma sustentável. A união de compreensão ecológica precisa e adequadas técnicas de manejo permitem uma exploração sustentável a longo-prazo dos recursos vegetais, assegurando manutenção de diversidade e provisões para o futuro.

A execução de estudos robustos para alcançar essa interface entre conservação e exploração demanda o uso de eficientes ferramentas analíticas. Dentre essas ferramentas, as linguagens de programação têm se sido importantes aliadas para obtenções de predições e resultados estatísticos confiáveis e informativos. A linguagem contida no software R é a mais amplamente utilizada para processamento de dados e análises de vegetação. O R engloba diversos pacotes importantes para análises de dados de plantas em diversos contextos ecológicos e agrários. Com seus diversos pacotes, o R permite a busca mais apurada pela compreensão de padrões e processos ecológicos, avaliação de impactos antrópicos sobre vegetação, monitoramentos e previsões de condições do solo para plantios e predições de efeitos de mudanças climáticas em florestas. Essa gama de possibilidades analíticas amplifica o acerto em tomadas de decisão com relação ao uso dos nossos recursos naturais de forma geral.

Este livro tem como objetivo trazer uma compilação de algumas potencialidades do software R para análise de vegetação, contribuindo para o aumento da capacidade técnica de diversos profissionais das áreas de Ciências da Terra ou Naturais no uso dessa poderosa ferramenta analítica. Para tal, os capítulos aqui presentes discorrem de forma aplicada sob temas em contextos ecológicos e agrários. Todos os capítulos possuem links de compartilhamento livre de dados e scripts com códigos para execução das análises que eles abordam no R. Assim, desejamos que o conteúdo aqui presente auxilie você leitor (a) em sua tarefa analítica, amplificando a obtenção de resultados informativos e potenciais de aplicação prática.

Écio Souza Diniz  
Pedro Manuel Villa



## SUMÁRIO

<b>CAPÍTULO 1</b> .....	<b>1</b>
BIOVEG – A PROTOCOL TO LEARN AND TEACH STATISTICS IN R USING VEGETATION DATA	
Écio Souza Diniz Jan Thiele	
<b>DOI 10.22533/at.ed.3552009031</b>	
<b>CAPÍTULO 2</b> .....	<b>11</b>
RAREFACTION AND EXTRAPOLATION OF SPECIES DIVERSITY DURING NEOTROPICAL FOREST SUCCESSION: AN R ROUTINE USING INEXT PACKAGE	
Pedro Manuel Villa Sebastião Venâncio Martins Écio Souza Diniz Antonio J. Pérez-Sánchez Gustavo Heringer Alice Cristina Rodrigues Daniela Schmitz Júnia Maria Lousada Herval Junior Pinto Andreza Viana Neri	
<b>DOI 10.22533/at.ed.3552009032</b>	
<b>CAPÍTULO 3</b> .....	<b>20</b>
PHYTOSOCIOLOGY IN R: A ROUTINE TO ESTIMATE PHYTOSOCIOLOGICAL PARAMETERS	
Gustavo Heringer Pedro Manuel Villa Andreza Viana Neri	
<b>DOI 10.22533/at.ed.3552009033</b>	
<b>CAPÍTULO 4</b> .....	<b>30</b>
ANÁLISE DE DADOS DE DESMATAMENTO COM R: VISUALIZAÇÃO INTERATIVA COM SHINY	
Carlos Eduardo Cardoso Mauricio Evandro Eloy João Paulo Martins dos Santos Alessandro Firmiano de Jesus	
<b>DOI 10.22533/at.ed.3552009034</b>	
<b>CAPÍTULO 5</b> .....	<b>43</b>
AVALIAÇÃO DE GRADIENTE PEDOAMBIENTAL USANDO ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) NA ANTÁRTICA MARÍTIMA	
Daniela Schmitz Pedro Manuel Villa Carlos Ernesto G.R. Schaefer Márcio Rocha Francelino	
<b>DOI 10.22533/at.ed.3552009035</b>	

<b>CAPÍTULO 6 .....</b>	<b>56</b>
DISTRIBUIÇÃO ESPACIAL DE FATORES AMBIENTAIS E ATRIBUTOS FLORESTAIS USANDO ROTINAS NO R	
Alice Cristina Rodrigues Pedro Manuel Villa Andreza Viana Neri	
<b>DOI 10.22533/at.ed.3552009036</b>	
<b>CAPÍTULO 7 .....</b>	<b>69</b>
SPATIAL RELATIONSHIP BETWEEN SOIL AND PHYTOSOCIOLOGICAL INDICATORS OF ECOLOGICAL RESTORATION IN AN ATLANTIC FOREST SITE	
Camila Santos da Silva Marcos Gervasio Pereira Rafael Coll Delgado Emanuel José Gomes de Araújo Cristiane Figueira da Silva Daniel Costa de Carvalho Shirlei Almeida Assunção Israel Oliveira Ramalho Deyvid Diego Carvalho Maranhão Ariovaldo Machado Fonseca Junior	
<b>DOI 10.22533/at.ed.3552009037</b>	
<b>CAPÍTULO 8 .....</b>	<b>82</b>
MODELAGEM ESPACIALIZADA DA EVAPOTRANSPIRAÇÃO REAL EM ÁREA DE REFLORESTAMENTO POR MEIO DO PACOTE AGRIWATER EM AMBIENTE R	
César de Oliveira Ferreira Silva Pedro Henrique Jandreice Magnoni	
<b>DOI 10.22533/at.ed.3552009038</b>	
<b>CAPÍTULO 9 .....</b>	<b>96</b>
IMPACTO DO FOGO NO BANCO DE SEMENTES DE FLORESTA ESTACIONAL SEMIDECIDUAL ALTOMONTANA NO QUADRILÁTERO FERRÍFERO, MG	
Júnia Maria Lousada Pedro Manuel Villa Gustavo Heringer Sebastião Venâncio Martins	
<b>DOI 10.22533/at.ed.3552009039</b>	
<b>CAPÍTULO 10 .....</b>	<b>110</b>
EFFECTS OF SPATIAL SCALE ON PEQUI ENTOMOFAUNA	
Gustavo Amorim Santos Deomar Plácido da Costa Suzana da Costa Santos Pedro Henrique Ferri	
<b>DOI 10.22533/at.ed.35520090310</b>	
<b>CAPÍTULO 11 .....</b>	<b>122</b>
PIPELINE DE EXPRESSÃO DIFERENCIAL EM R APLICADO À <i>Arabidopsis thaliana</i>	
Sheila Tiemi Nagamatsu Lucas Miguel de Carvalho	



Luciana Souto Mofatto  
Nicholas Vinícius Silva  
Marcelo Falsarella Carazzolle  
Gonçalo Amarante Guimarães Pereira

**DOI 10.22533/at.ed.35520090311**

**CAPÍTULO 12 ..... 138**

**MODELAGEM DE CRESCIMENTO DE CANA-DE-AÇÚCAR E CANA ENERGIA SOB O ESTÍMULO DE REGULADOR DE CRESCIMENTO**

Luís Guilherme Furlan de Abreu  
Lucas Miguel de Carvalho  
Maria Carolina de Barros Grassi  
Gonçalo Amarante Guimarães Pereira

**DOI 10.22533/at.ed.35520090312**

**CAPÍTULO 13 ..... 150**

**INFLUÊNCIA DA SUPLEMENTAÇÃO POR FLAVONOIDE NO CRESCIMENTO DE CLONES COMERCIAIS DE *E. urophylla* e *E. urograndis***

Nicholas Vinícius Silva  
Luciana Souto Mofatto  
Mariana Teixeira Rebouças  
Lucas Miguel de Carvalho  
Sheila Tiemi Nagamatsu  
Marcelo Falsarella Carazzolle  
Jorge Lepikson Neto  
Gonçalo Amarante Guimarães Pereira

**DOI 10.22533/at.ed.35520090313**

**SOBRE OS ORGANIZADORES..... 166**

**ÍNDICE REMISSIVO ..... 167**

## BIOVEG – A PROTOCOL TO LEARN AND TEACH STATISTICS IN R USING VEGETATION DATA

*Data de aceite: 12/02/2020*

### Écio Souza Diniz

Laboratório de Geoprocessamento e Sensoriamento Remoto, Departamento de Engenharia Florestal, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil.  
eciordiniz@gmail.com

### Jan Thiele

Thünen Institute for Biodiversity, Federal Research Institute for Rural Areas, Forestry and Fisheries, 38116, Braunschweig, Niedersachsen, Germany.  
jan.thiele@thuenen.de

**ABSTRACT:** The wide range of vegetation types that exist throughout the world, and particularly tropical rainforests, are crucial for the provision of ecosystem goods and services. Efficient and robust analytical tools are required to extract valuable information from vegetation data, thereby enhancing our comprehension of the interaction of plants with the environment in which they occur. Such information is increasingly valuable in the context of rapid global change. The growing use of programming languages such as R, in computational statistics, has facilitated the development of ever more complex analyses, thus delivering significant advances in our understanding of vegetation

dynamics. In this chapter, we present Bioveg, a freely available set of R scripts and datasets for the teaching and learning of statistical analysis using R. We use vegetation data to illustrate a range of statistical analyses, from commonly used basic statistics to many more advanced and complex procedures. The data we explore in this text was collected from temperate Vegetation in Germany and Brazilian Atlantic Forests. The datasets are freely available to download. We hope to contribute to the distribution of understanding of the proper use of statistical analyses in R, and their application for vegetation analysis.

**KEYWORDS:** R language, Statistics, Vegetation Data, Computational Programming

### BIOVEG - UM PROTOCOLO PARA APRENDER E ENSINAR ESTATÍSTICAS EM R COM DADOS DE VEGETAÇÃO

**RESUMO:** A grande variedade de tipos de vegetação em todo o mundo, destacando as florestas tropicais, é crucial no fornecimento de bens e serviços ecossistêmicos. Devido à sua importância, os dados da vegetação exigem ferramentas analíticas eficientes e robustas para extrair informações valiosas, capazes de melhorar nossa compreensão da interação das plantas com o ambiente em que ocorrem, especialmente em tempos de



rápidas mudanças globais. O crescente uso de estatísticas computacionais por meio de linguagens de programação possibilitou alcançar níveis mais altos de precisão na aquisição de tais conhecimentos, o que destaca o uso do software R. Neste capítulo, apresentamos o Bioveg, um conjunto gratuito de scripts para R e conjuntos de dados para aprender e ensinar estatística em R usando dados de vegetação para exemplificar uma gama significativa de análises estatísticas distintas, das básicas às avançadas. Os conjuntos de dados fornecidos abrangem dados registrados pelos autores na Vegetação Temperada Alemã e Florestas Atlânticas Brasileiras e podem ser baixados gratuitamente. Esperamos contribuir para a difusão do conhecimento geral sobre estatística em R e sua aplicação na análise de vegetação.

**PALAVRAS-CHAVE:** Linguagem R, Estatística, Dados da Vegetação, Programação Computacional

## 1 | INTRODUCTION

Globally, all types of vegetation, and particularly tropical rainforests, are essential providers of ecosystem goods and services (CHAZDON, 2008; CHAZDON et al., 2009) such as carbon sequestration (PAN et al., 2011), protection of water supplies and flood control (ELLISON; FUTTER; BISHOP, 2012), soil maintenance (DITT et al., 2010) and many others. In times of rapid environmental change, understanding the interactions of plants with the ecosystem in which they occur is crucial. It is of particular importance to understand the likely consequences of environmental change on ecosystem dynamics, in order to generate plans to mitigate the negative impacts of environmental change on the provision of ecological goods and services. Modern computational tools are fundamental to delivering significant advances in this field.

The R environment is an integrated suite of software facilities for the manipulation of data using R programming language. Within the last two decades it has become one of the most commonly used programming languages to process data and execute a huge range of statistical analyses, for both plants and other organisms (CRAWLEY, 2012). There is a vast range of packages which have been developed for numerous statistical approaches, including many analytical specificities for multiple organisms, plant ecology and evolution (e.g. BORCARD; GILLET; LEGENDRE, 2011; CASSINI, 2013; GARAMSZEGLI, 2014; SWENSON, 2014; PARADIS; CLAUDE; STRIMMER, 2004). The constant development of R packages and functions has facilitated important enhancements in our comprehension of plant communities (e.g. BARTON, 2016; BATES et al., 2015; FOX; WEISBERG, 2010; JIN; QIAN, 2019; KECK et al., 2016; KEMBEL et al., 2010; PARADIS; CLAUDE; STRIMMER, 2004; PINHEIRO et al., 2017; RITZ; STREIBIG, 2008).

Despite the significant benefits of using R and its associated packages for data processing and statistical analyses, there are regions (e.g. the tropics) where its use

remains few explored. We present this text in order to contribute to the dissemination of knowledge of the use of R for vegetation data analyses. Bioveg is a freely available set of R scripts and data, with examples for a wide range of statistical analyses, covering basic statistical analyses and more advanced statistical techniques. The Bioveg content comes with real data from Brazilian Atlantic Neotropical Forests and German Temperate Vegetation. This data forms the basis of the examples used to explain each type of analysis, and can be used to practice running analyses in R.

## 2 | DATA DETAILS

The data provided in the Bioveg dataset file are real data collected in the field and provided by the authors. The original data was randomized in order to avoid overlapping with published data or creating a conflict of interest. The datasets regarding Brazilian Atlantic Forest encompass three distinct forest types; semideciduous, Ombrophilous (rain) and gallery forest. Data collection took place between 2011 and 2018, using a standard sampling design of plots and subplots (10 per 10 m), at 3 sites in Southeast and Northeast Brazil.

The geographical coordinates of each study site were also altered to prevent overlap with published work. For each subplot within study sites, the diameter at breast height (DBH) and basal of each tree  $\geq 10$  and 15 cm were measured. Abundance and richness of each species within each subplot were also registered. Data collection was repeated after four and five years (depending on the study site location) to enable the annual rates of mortality, recruitment, gain and loss of basal area, for tree individuals within the subplots to be calculated. Climatic data (mean precipitation and temperature) were retrieved from WorldClim and extracted to each site using ArcMap from ArcGIS. Further environmental data regarding soil parameters (pH, P, K, Ca, Mg and Al), and species and vegetation types matrices are provided, to demonstrate how to perform multivariate analyses.

For the German Temperate Vegetation datasets, the field sampling took place in sites invaded by the species Giant Hogweed (*Heracleum mantegazzianum* Sommier & Levier), a monocarpic perennial herbaceous flowering plant in the family Apiaceae that is a real problem in Europe (BOOY et al., 2005) and North America (PYSEK et al., 2007) due to its aggressively invasive nature and the phytotoxic chemicals found in its sap. The data was recorded between 2002 and 2003 in South and West Germany. The sampling design was based on the Braun-Blanquet's Relevés method. The subplots were classified into broad habitat categories (Habitat type) based on physiognomy, type of disturbance and land use.

The habitat categories were agricultural grassland, which includes meadows and

managed pastures; ruderal grassland, referring to meadows or pastures abandoned or irregularly maintained; tall herb vegetation, which includes several nitrophilous tall herb communities (of the Galio-Urticetea class, perennial or biannual vegetation types of mesic to wet habitats); woodland, for areas with more or less closed tree layer; and wasteland, which refers to areas of open vegetation that do not belong to either pasture or tall grass communities. Land use was classified in grassland, with regular cutting or grazing; maintenance, cut once or twice a year to prevent shrub invasion; 'none', no signs of cutting, grazing or other anthropic uses of the land. Disturbance types considered were deposition of organic materials ('deporg'), flooding by rivers ('flood'), 'none', disturbance of the topsoil ('soildis') and removal of trees or shrubs ('treeshrub').

For each study area, five soil samples were collected at a depth of 20 cm and homogenized and sieved, which was used to determine the content of major plant nutrients: total nitrogen (N-Perc), plant phosphorus availability in mg/100g (P\_mg\_100g) and potassium in mg/100g (K\_mg\_100g). Further, the altitudes of the study areas were obtained using GPS, and inclination and exposure were obtained with the use of inclinometer and compass. Additionally, height and percentage of Giant Hogweed cover were measured.

Overall, the content of Bioveg includes six dataset files: "Bioveg3.csv" containing most of the data on Brazilian Atlantic Forest, supported by the matrices "matrixspp.txt", "matrixtype.txt" and "matrixenv.txt" that relate to the tree community composition, type of vegetation and the soil parameters of these communities, respectively. Further, the file "Logit.csv" includes binary data (success or failure) describing tree recruitment. Finally, the file "Hogweed.csv" contains the complete German Temperate Vegetation dataset.

### 3 | ANALYSIS CONTENT

The examples in this text are all performed using the data provided in the Bioveg data file (6 datasets). The function and application of each of the analyses, and the particular relevance of these techniques in vegetation analysis, is explained. A complete list of the statistical analyses covered in Bioveg can be found in Table 1. Many of the examples can be executed using built-in functions in R, but some require prior installation of specific packages (Table 1).

ANALYSIS TYPE	STATISTICS APPLIED	R PACKAGES
<b>Normality test</b>	Shapiro Wilk test and residuals	Built-in Function
<b>Data Transformation</b>	Log, square, inverse and Arcsine square root	Built-in Function
<b>One sample test</b>	<i>Parametric</i> : T-test; <i>Nonparametric</i> : Wilcoxon Signed Rank Test	Built-in Function
<b>Two sample tests</b>	<i>Independent samples</i> : Nonparametric: Wilcoxon Rank-Sum Test (Mann-Whitney U-test); <i>Parametric</i> : T-test; <i>Dependent sample</i> : Wilcoxon Rank-Sum; Paired T-test	Built-in Function
<b>Analyses of variance</b>	<i>Parametric</i> : ANOVA One Way and Tukey post-hoc; <i>Nonparametric</i> : Kruskal-Wallis and Dunn's post-hoc test	Built-in Function and dunn.test
<b>Independence test</b>	Chi-squared	Built-in Function and MASS
<b>Correlation analysis</b>	Pearson, Spearman and Variance Inflation Factor (VIF)	Built-in Function and PerformanceAnalytics
<b>Regression modeling</b>	<i>Linear</i> : LM (Linear Model), GLM (Generalized Linear Models, GLMM (Generalized linear mixed models), Logistic regression, LME (Linear mixed models), GLS (Generalized Least squares) and post-hoc Tukey HSD and LmerTest tests. <i>Nonlinear</i> : Quadratic Model, Generalized Non-Linear Models (GNM), GAM (Generalized Additive Models) and GAMM (Generalized Additive Mixed Models). Note: There are examples treating spatial and temporal correlation.	Lme4, nlme, gnm, mgcv, multcomp, lmerTest, MASS, car, gmodels and ROCR
<b>Model selection</b>	Akaike information criterion (AIC) and AIC model average	MuMIn
<b>Linearity diagnostic</b>	Based on residuals of linear regression models	car
<b>Spatial autocorrelation</b>	Based on Moran's I statistics applied on linear regression models	lctools, ncf, nlme,
<b>Supervised learning</b>	Random Forest	randomForest
<b>Multivariate analysis</b>	NMDS, ANOSIM, CCA, PCA, PERMANOVA and INDVAL	BiodiversityR, vegan, indicpecies, permute and lattice

Table 1: Analyses, statistical methods and their related packages present in Bioveg. Built-in Function means package independent functions.



Regression analysis is used to examine relationships between variables in a wide range of scientific areas. In regression modelling we test how accurately the response variable is predicted by the explanatory variable. This type of statistical analysis is by far the most commonly used analysis in vegetation science and a large part of Bioveg is allocated to explain the various types of regression model (linear and non-linear) that can be used. Further detail and examples about some of these models can be found in Diniz and Thiele (2018). We focus on generalized models (fixed and mixed effects, GLM and GLMM) given the multitude of applications they offer and their ability to account for the variety of error distributions that are likely to be encountered in the dependent variable (e.g. Poisson, Binomial, Gaussian, Gamma).

It is vital when performing regression analysis that a number of criteria are met, violation of these criteria undermines the validity of the conclusions which are drawn from the model. Therefore, a number of important pre-tests must be conducted when constructing a model are explained in this text: a test for normality of the distribution; linearity diagnostics to verify whether there is a linear relationship between the predictor and dependent variables; correlation tests to check for collinearity among predictors and removal of variables that are found to highly correlate. We also describe how the model selection approach (Akaike Criterion) can be used to select the most useful and informative model from a global model which includes all possible predictor variables

We also provide examples of the use of smoothing functions for cases of generalized models which contain non-linear relationships (i.e. GAM, GAMM and GNM) between predictor and dependent variables. Lastly, we introduce the topic of machine learning by means of the decision tree approach using Random Forest. Random Forest is one of the most robust approaches of supervised machine learning used in predictive modelling; we therefore provide an example of this technique.

## 4 | DEMONSTRATIVE EXAMPLE

### 4.1 Predicting species richness with Random Forest

The algorithms of Random Forest constitute an ensemble learning method allowing for both classification and regression trees (BREIMAN, 2001). That is, these algorithms construct a multitude of decision trees during training and output the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (BREIMAN, 2001; HO, 1998). The decision tree is a type of modeling that operates with information gain on each node, classifying data points with greater information increment on each node (DIETTERICH, 2000). When the information gain ratings for all nodes are depleted, the model achieves its optimal performance result. Thus, the final decision is made, based on the majority, i.e. average, of the

trees and is chosen by the random forest (DIETTERICH, 2000). In Random Forest, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error, since this is internally estimated using the out-of-bag (OBB) error for each data point recorded and averaged over the forest (JAMES et al., 2013). The averaging of the difference in out-of-bag error before and after the permutation over all trees provides an importance score for the predictors, i.e., which are more or which are less contributive to the overall model performance (HO, 1998). Moreover, the adjusted decision trees in Random Forest enhance its efficiency in dealing with overfitting in the training set data (BREIMAN, 2001).

In our example for Random Forest in the R script of Bioveg, we want to predict species richness from the Hogweed dataset on German Vegetation, utilizing all the other variables (16) as predictors. Thus, we used the function `randomForest` of the package `randomForest` (LIAW; WIENER, 2002) and carried out a regression in the training set (70% of the data). For further detail on this modeling technique, its steps and codes for reproduction of this example, please, refer to the Random Forest section in our freely available script. Figure 1 below shows in descending order the importance of each of the 16 predictors for the model performance, in predicting species richness.

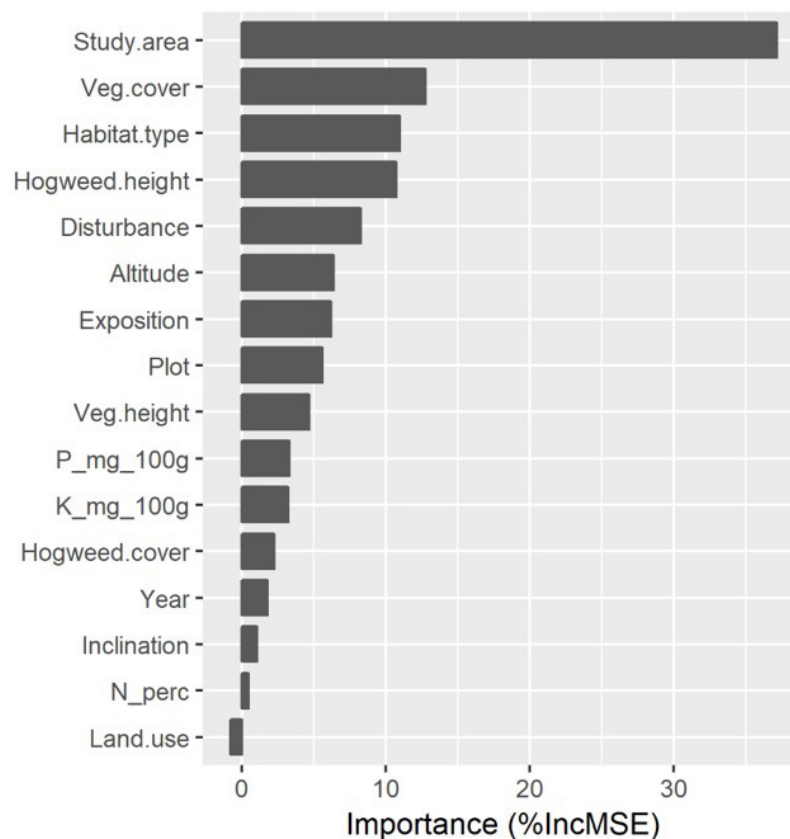


Figure 1: Variable Importance of predictors for species richness from a Random Forest regression model.

In accordance with the output of the Figure 1, study area, vegetation cover (Veg.

cover), vegetation type (Habitat.type) and hogweed cover are the most important variables for the model performance in predicting species richness, while from disturbance the importance starts to decline to all other predictors. The level of importance from potassium (K\_mg\_100g) to land use could be considered negligible for the model predictive power, i.e. withdrawing these variables from the model would have little effect on the performance of the model.

Bioveg details each analysis and tests it contains, providing a theoretical and practical overview of application for the user. The syntax and code provided for each analysis may be adapted by the user to apply, test or exercise with their own data.

## 5 | ACCESSIBILITY TO BIOVEG

The whole content of Bioveg (datasets and script) can be freely accessed and download in the links:

Research Gate: <http://doi.org/10.13140/RG.2.2.12488.26889>

GitHub: <https://github.com/eciodiniz/Bioveg>

With the DOI (10.13140/RG.2.2.12488.26889) provided by license of Research Gate, Bioveg can be citable, please where possible support and give credit to this initiative by citing it in your studies.

## 6 | FINAL CONSIDERATIONS

Statistics and computer programming are central skills in the 21st century, and further progress in this field is vital if we are to address the increasingly complex problems that are arising in the modern world. Among these problems, environmental degradation and biodiversity loss demand urgent solutions. Developments in statistical analysis software are promising huge potential gains in the fundamental knowledge of the relationships and mechanisms maintaining ecosystem health. With greater knowledge we can begin to develop interventions to address these problems, but professionals need training in order to exploit the opportunities these new techniques offer. Our intention is that Bioveg will be used by professional and trainee researchers in the relevant fields (e.g. biologists, ecologists and foresters), to develop their use of R as a venue for data analysis and take advantage of the many benefits this platform can provide in the analysis of vegetation data.

## REFERENCES

BARTON, K. **MuMIn: Multi-Model Inference.**, 2016. Disponível em: <<https://cran.r-project.org/package=MuMIn>>

- BATES, D. et al. Fitting Linear Mixed-Effects Models using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 51, 2015.
- BOOY, O. et al. **The giant hogweed best practice manual. Guidelines for the management and control of an invasive alien weed in Europe**. Hoersholm: Forest and Landscape Denmark, 2005.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- CHAZDON, R. L. Beyond deforestation: restoring forests and ecosystem services on degraded lands. **science**, v. 320, n. 5882, p. 1458–1460, 2008.
- CHAZDON, R. L. et al. The potential for species conservation in tropical secondary forests. **Conserv Biology**, v. 23, n. 6, p. 1406–1417, 2009.
- CRAWLEY, M. J. **The R book, 2nd Edition**. Chichester: Wiley, 2012.
- DIETTERICH, T. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. **Machine Learning**, v. 40, n. 2, p. 139–57, 2000.
- DINIZ, E. S.; THIELE, J. **Modelagem de regressão no R: aplicações em análises de vegetação**. Saarbrücken: Novas Edições Acadêmicas, 2018.
- DITT, E. H. et al. Forest Conversion and Provision of Ecosystem Services in the Brazilian Atlantic Forest. **Land Degradation & Development**, v. 21, n. 6, p. 591–603, 2010.
- ELLISON, D.; FUTTER, M. N.; BISHOP, K. On the forest cover–water yield debate: from demand- to supply-side thinking. **Global Change Biology**, v. 18, p. 806–820, 2012.
- FOX, J.; WEISBERG, S. **An R companion to applied regression**. Thousand Oaks: Sage, 2010.
- HO, T. K. The Random Subspace Method for Constructing Decision Forests. **EEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 8, p. 832–844, 1998.
- JAMES, G. et al. **An Introduction to Statistical Learning**. New York: Springer, 2013.
- JIN, Y.; QIAN, H. V. PhyloMaker: an R package that can generate very large phylogenies for vascular plants. **Ecography**, v. 42, n. 8, p. 1353–1359, 2019.
- KECK, F. et al. phylosignal: an R package to measure, test, and explore the phylogenetic signal. **Ecology and Evolution**, v. 6, n. 9, p. 2774–2780, 2016.
- KEMBEL, S. W. et al. Picante: R tools for integrating phylogenies and ecology. **Bioinformatics**, v. 26, n. 11, p. 1463–1464, 2010.
- LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2, n. 3, p. 18–22, 2002.
- PAN, Y. et al. A large and persistent carbon sink in the world's forests. **Science**, v. 333, n. 6045, p. 988–993, 2011.
- PARADIS, E.; CLAUDE, J.; STRIMMER, K. APE: analyses of phylogenetics and evolution in R language. **Bioinformatics**, v. 20, n. 2, p. 289–290, 2004.
- PINHEIRO, J. et al. **nlme: Linear and Nonlinear Mixed Effects Models\_ . R package version 3.1-**



131, 2017. Disponível em: <<https://cran.r-project.org/package=nlme>>

PYSEK, P. et al. **Ecology and management of giant hogweed (*Heracleum mantegazzianum*)**. [s.l.] CABI, 2007.

RITZ, C.; STREIBIG, J. C. **Nonlinear Regression with R**. New York: Springer, 2008.

## RAREFACTION AND EXTRAPOLATION OF SPECIES DIVERSITY DURING NEOTROPICAL FOREST SUCCESSION: AN R ROUTINE USING INEXT PACKAGE

Data de aceite: 12/02/2020

### **Pedro Manuel Villa**

Laboratório de Restauração Florestal,  
Departamento de Engenharia Florestal,  
Universidade Federal de Viçosa, Viçosa, Minas  
Gerais, Brazil.  
villautana@gmail.com

### **Sebastião Venâncio Martins**

Laboratório de Restauração Florestal,  
Departamento de Engenharia Florestal,  
Universidade Federal de Viçosa, Viçosa, Minas  
Gerais, Brazil.

### **Écio Souza Diniz**

Departamento de Engenharia Florestal,  
Universidade Federal de Viçosa, Viçosa, Minas  
Gerais, Brazil

### **Antonio J. Pérez-Sánchez**

Thünen Institute of Biodiversity, Bundesallee 65,  
38116 Braunschweig, Germany/ Biodiversity of  
Agricultural Landscapes, Institute of Geoecology,  
Technische Universität Braunschweig,  
Braunschweig, Germany

### **Gustavo Heringer**

Programa de Pós-Graduação em Ecologia  
Aplicada, Universidade Federal de Lavras - UFLA,  
Minas Gerais, Brasil.

### **Alice Cristina Rodrigues**

Programa de pós-graduação em Botânica,  
Departamento de Biologia Vegetal, Universidade  
Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

### **Daniela Schmitz**

Programa de pós-graduação em Botânica,  
Departamento de Biologia Vegetal, Universidade  
Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

### **Júnia Maria Lousada**

Programa de pós-graduação em Botânica,  
Departamento de Biologia Vegetal, Universidade  
Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

### **Herval Junior Pinto**

Programa de pós-graduação em Botânica,  
Departamento de Biologia Vegetal, Universidade  
Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

### **Andreza Viana Neri**

Programa de pós-graduação em Botânica,  
Departamento de Biologia Vegetal, Universidade  
Federal de Viçosa, Viçosa, Minas Gerais, Brazil.

**ABSTRACT:** The Amazon basin harbors the most important tropical forest in the world. Thus, quantifying and assessing changes in biological diversity are central aspects for Amazon forest management and conservation. Hill numbers approach is increasingly used to characterize the forest diversity. Here, we evaluate changes in tree species diversity along secondary succession by using the rarefaction and extrapolation (Hill numbers) approach in the Northern Amazon forest. We selected three sites over a 12 km<sup>2</sup> extension of an old-growth forest matrix with secondary forest patches of different stand ages. At each site, plant diversity of trees having diameter  $\geq 5$  cm was assessed.

We analyzed differences in woody species diversity among the four successional stages and the old-growth forest using individual-based rarefaction and extrapolation and coverage-based approaches. The lowest evenness was observed in the forest at the earliest successional stage, due to the presence of a few dominant species in the community; however, stabilized at the subsequent successional stages. We conclude that the forest stand heterogeneity between successional stages can be analyzed by the differences in diversity using individual-based rarefaction and extrapolation and coverage-based approaches.

**KEYWORDS:** bootstrapping, forest regrowing, sampled-based rarefaction and extrapolation, second-growth forests; tree dominance.

## RAREFAÇÃO E EXTRAPOLAÇÃO DA DIVERSIDADE DE ESPÉCIES DURANTE SUCESSÃO DE FLORESTA NEOTROPICAL: UMA ROTINA R USANDO O PACOTE INEX

**RESUMO:** A bacia amazônica abriga a floresta tropical mais importante do mundo. Assim, quantificar e avaliar mudanças na diversidade biológica são aspectos centrais para o manejo e conservação da floresta amazônica. A abordagem dos números de Hill é cada vez mais usada para caracterizar a diversidade da floresta. Aqui, avaliamos mudanças na diversidade de espécies arbóreas ao longo da sucessão secundária, usando a abordagem de rarefação e extrapolação (números de Hill) na floresta do norte da Amazônia. Selecionamos três locais ao longo de uma extensão de 12 km<sup>2</sup> de uma matriz florestal madura com fragmentos de florestas secundárias de diferentes idades. Em cada localidade, foi avaliada a diversidade de plantas com diâmetro  $\geq 5$  cm. Analisamos as diferenças na diversidade de espécies lenhosas entre quatro estágios sucessionais e a floresta de crescimento antigo, usando abordagens de rarefação e extrapolação baseadas em indivíduos e cobertura baseadas em amostras. A menor equabilidade foi observada na floresta no estágio inicial de sucessão, devido à presença de algumas espécies dominantes na comunidade; no entanto, estabilizou nos estágios sucessionais subsequentes. A heterogeneidade ambiental entre estágios sucessionais pode ser analisada pelas diferenças de diversidade, utilizando abordagens baseadas em rarefação e extrapolação baseada em indivíduos e cobertura baseadas em amostras.

**PALAVRAS-CHAVE:** bootstrapping, dominância de árvores, floresta secundária, rarefação e extrapolação com base em amostras, regeneração florestal

### 1 | INTRODUCTION

The Amazon region provides important ecosystem services that have local and global impact (Zhang et al., 2015). However, human-induced biodiversity change impairs ecosystem services crucial to human well-being. For example, cutting and burning agriculture practiced for centuries are still the main agricultural system that

sustains human populations in the Amazon (Arroyo-Kalin et al., 2012; Villa et al., 2017). Intensification by shifting cultivation (SC) can seriously threaten the old-growth forests (OGF) and their biodiversity at the local scale (Jakovac et al., 2015; Barlow et al., 2016; Villa et al., 2018a). However, second-growth forest (SGF) recovery after anthropogenic disturbances represents an important reservoir of biodiversity (Chazdon, 2014; Rozendaal et al., 2019). Thus, information about species diversity recovery during succession can help to enhance our understanding of SGF dynamics and improve predictions on the impact of future climate and land-use changes for effective forest restoration and conservation policies.

Forest regeneration based in ecological succession approach, consists of the recovery of different components of the ecosystem, such as diversity, structure, and function over the time (Chazdon, 2014; Villa et al., 2018b; Poorter et al., 2019; Rozendaal et al., 2019). Most of the previous studies on the secondary succession in tropical forests have addressed the changes in species diversity comparing SGF and OGF (Chazdon et al., 2014; Rodendaal et al., 2019). For instance, this might be achieved by using the rarefaction approach (i.e. sampled and individual-based) with Hill numbers, species richness ( $q = 0$ ), exponential of Shannon entropy (Shannon's diversity,  $q = 1$ ), and inverse Simpson concentration (Simpson diversity,  $q = 2$ ) between reference samples from different assemblages using datasets from old-growth and second-growth forests (Colwell et al., 2012; Chao et al., 2014). An advantage of Hill numbers is that they are all expressed in units of effective numbers of species: the number of equally abundant species that would be needed to give the same value of a diversity measure (Chao et al., 2014).

This approach of Hill numbers can allow evaluating the balance between dominance and evenness in the tree community during the succession (Colwell et al., 2012; Chao et al., 2014). Thus, Integrated curves based on sampling theory that smoothly link rarefaction (interpolation) and prediction (extrapolation) standardize samples on the basis of sample size or sample completeness and facilitate the comparison of biodiversity data along succession (Chao et al., 2014). On the other hand, there is an alternative approach that defines the non-asymptotic Hill numbers as the average Hill numbers over many samples of size  $m$  taken from the entire assemblage. This integrated sample-size- or coverage-based rarefaction/extrapolation sampling curve represents the estimated diversity accumulation curve based on the reference sample (Chao et al., 2014).

In this study, we analyzed the changes in tree species diversity along secondary succession using the individual-based rarefaction and extrapolation (Hill numbers) and coverage-based approaches in the Northern Amazon forest. In order to evaluate the tree species diversity patterns (i.e. three Hill numbers), we asked the following scientific questions: does species diversity change during SGF recovery in comparison



with an old-growth forest? We hypothesize that tree diversity after shifting cultivation should be lower in the early successional assemblage, that is dominated by the abundance of one or a few common species (i.e. using individual-based rarefaction and extrapolation curves), compared with the old-growth forests. Thus, we presumed that areas undergoing regeneration may harbor higher tree species diversity during succession where dominance declines and increases the species coexistence of early successional and fast-growing pioneer species competitive for light.

## 2 | MATERIAL AND METHODS

### 2.1 Study area

The study was conducted in a semi-deciduous forest on the Cataniapo river basin, around and close to the indigenous communities of Piaroa de Gavilán and Sardi (613020 E, 679290 N, UTM), at the municipality of Atures, Amazon state, Venezuela. Climate in the study area is tropical rainy with dry season of two-months (December and January). Mean annual temperature and annual precipitation are 28 °C and 2700 mm, respectively. The predominant soil types are Oxisols (Latosols) and Ultisols (red clay soils), with low cation exchange capacity, low nutrient content and high acidity levels (Villa et al., 2018a).

### 2.2 Forest survey

We selected three sites over a 12 km<sup>2</sup> extension of land in an old-growth forest matrix with secondary-forest patches, of different stand ages, which have been abandoned after a single cycle of traditional shifting cultivation and left to regenerate naturally. Sampling sites were systematically selected according to their successional stages, being identified with the assistance of local farmers and experts. A total 45 plots (20 x 50 m = 1000 m<sup>2</sup>, each) were established from January 2009 to December 2012. At each site, we selected four secondary forests at different successional stages: 5, 10, 15, and 20 years of succession; and one old-growth forest (> 100 years old). Three plots were established within each secondary forest as well as within the old-growth forest, at each site.

In each plot, all trees having diameter at breast height (DBH)  $\geq$  5 cm were identified to the species level and tagged for measurement. The tagged specimens were identified at the Julian Steyermark herbarium of Puerto Ayacucho (Ministry of the Environment, Amazonas state, Venezuela). The *Angiosperm Phylogeny Group IV* (APG IV, 2016) was used to classify taxa.

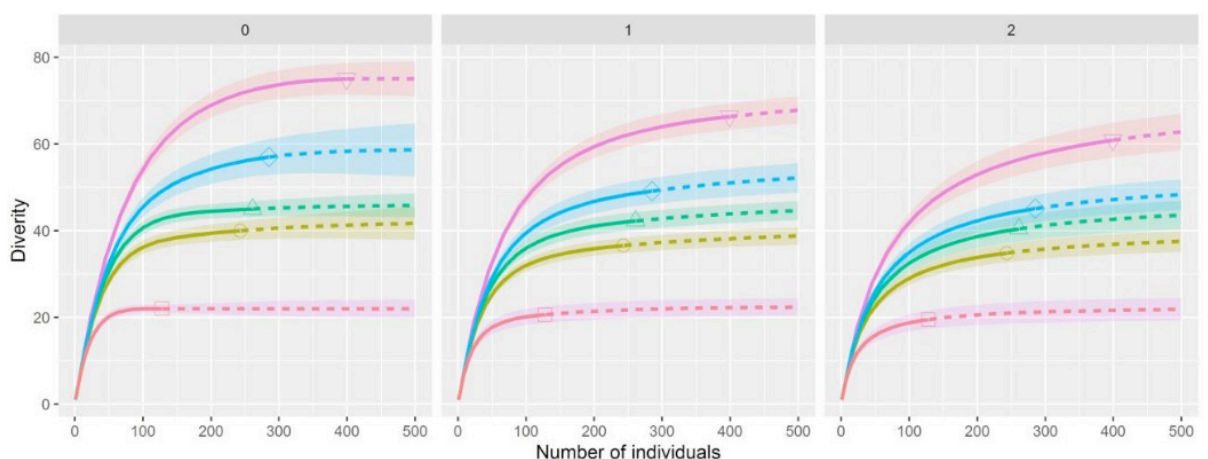
## 2.3 Statistical analyses

All analyses were carried out using R version 3.4.3 (R Core Team 2018). We analyzed differences in species richness and diversity among the four second growth (SGF) forests with different successional stages and the old-growth forest using both individual-based (Colwell et al., 2012) and coverage-based (Chao and Jost, 2012) approaches. Species diversity curves were constructed with Hill numbers: species richness ( $q = 0$ ), exponential of Shannon entropy (Shannon's diversity,  $q = 1$ ), and inverse Simpson concentration (Simpson diversity,  $q = 2$ ), for both individual-based rarefaction and extrapolation curves (Jost, 2007; Colwell et al., 2012; Chao et al., 2014). We then constructed coverage-based rarefaction and extrapolation curves of species richness for each assemblage. Extrapolations were made from a presence/absence data matrix (Hill numbers of order 0), being higher than thrice the sample size (Colwell et al., 2012). Sample-based rarefaction/extrapolations with 95% confidence intervals were computed using the 'iNEXT' package (Hsieh et al., 2016).

All three Hill numbers were estimated as the mean of 100 replicate bootstrapping runs to estimate 95% confidence intervals. Whenever the 95% confidence intervals did not overlap, species numbers differed significantly at  $p < 0.05$  (Colwell et al., 2012).

## 3 | RESULTS

We observed differences in species diversity among the successional stages. The species richness curves showed a clear trend of change over time after shifting cultivation (Fig. 1). Individual-based rarefaction and extrapolation curves showed higher tree species richness in the old-growth forest than in SGF at early and intermediary successional stages. Although their confidence intervals overlapped slightly only obtained through extrapolation among 15 and 20 years of succession after shifting cultivation (Fig. 1). Effective number of species calculated by the Hill number  $q = 0$  reach an asymptote after extrapolated.



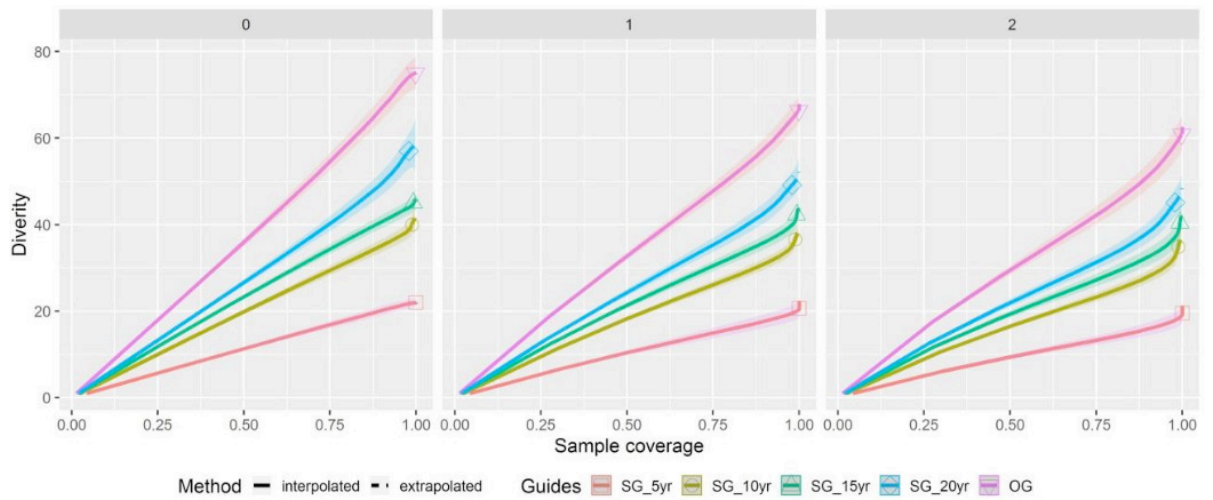


Fig. 1. Individual-based and coverage-based rarefaction (solid line) and extrapolation curves (dashed lines) of tree diversity on the Hill numbers ( $q = 0, 1, 2$ ) for the four successional stages and old growth forest.

All three Hill numbers increased consistently along the succession to old-growth forest. However, there was a decrease in the diversity values among successional stages ( $q = 0 > q = 1 > q = 2$ ). Shannon's and Simpson's diversities were asymptotic, with no increase in values even when they were extrapolated to double the reference sample size, except for the forest at the SGF earliest successional stage (Fig. 1).

Coverage-based rarefaction and extrapolation curves revealed higher species richness in the old-growth forest than in forests at early and intermediary successional stages. Sample coverage for all sites remained under  $\sim 100\%$  (Fig. 1).

#### 4 | DISCUSSION

The effective number of species richness calculated by the Hill numbers ( $q = 0$ ) reached an asymptote after being extrapolated. Its confidence intervals slightly overlapped when obtained through extrapolation, in secondary forests with 15 and 20 years of succession after shifting cultivation (Fig. 1). As for the rarefaction curves constructed with the other Hill numbers  $q = 1$  and  $q = 2$  (exponential of Shannon and inverse Simpson, respectively), they also indicated that tree diversity was significantly higher in OGF than in secondary forests. However, the estimated diversity with these numbers tended to decrease when compared with richness ( $q = 2 < q = 1 < q = 0$ ), except during the earliest successional stage. This pattern provides an additional evidence that corroborates the facts that there is less evenness during early successional stages and that dominance tends to decrease along succession. The increase in diversity from early succession through two decades later may also be related to the decreased dominance of species with highest abundance values. Similarly, the decrease observed with inverse Simpson, which estimates the diversity

of dominant species (Chao et al., 2014), can be partly explained by a more uniform distribution of abundance. On the other hand, coverage-based rarefaction curves showed that the observed number of individuals (i.e. up to the points indicated by the shapes in each forest type) indicate 90% of the individuals in the community belonging to species in the samples.

#### 4.4 Implications for forest management and conservation

Understanding how shifting cultivation affects the long-term recovery of tropical secondary forests is critical to assess the conservation value of biodiversity. While there is a constant pressure for land-use change on Amazonian forests, their future will also depend on the improvement of conservation, management and restoration methods. Our results demonstrate the high potential for diversity recovery of those forests after they have undergone agricultural disturbance, as their levels of species richness can approach the ones of old-growth forests within two decades (Villa et al. 2018b; Rozendaal et al., 2019). However, it is also necessary to consider the importance of species composition recovery, which is a much slower process (Finegan, 1996; Rozendaal et al., 2019), that may be relevant to the multiple ecosystem functions. Thus, it should be taken as a premise that a single cycle of shifting cultivation in the northern region of the Amazon still represents an opportune time to take conservation actions. Therefore, it would be necessary to avoid a second or further cycles of shifting cultivation in the same previously used areas. Otherwise, such land use pattern based on multiple cycles of shifting cultivation can reduce diversity recovery and consequently expand the proportion of degraded areas in the Amazon forest.

Despite the recent research findings that diversity in tropical forests is made up by different components (i.e. taxonomic, functional, and phylogenetic diversities), taxonomic diversity still remains with a high relevance for the understanding of ecological patterns and processes that take place during succession, especially due to its importance for recovery practices. On the other hand, it may be relevant to assess other taxonomic diversity metrics that could also contribute to elucidate emergent patterns during succession. The use of Hill numbers for instance has been few explored in the studies on secondary forests succession. For that reason, we propose the adoption of the individual-based rarefaction and extrapolation (Hill numbers) and coverage-based as complementary approaches to evaluate the recovery potential of forests after anthropogenic disturbances.

## 5 | CONCLUSIONS

The diversities estimated with Hill numbers  $q = 1$  and  $q = 2$  lead us to conclude



that they tend to decrease when compared with richness ( $q = 0$ ), except during the earliest successional stage. This pattern corroborates the lower evenness observed during early stages and the decreasing dominance observed along succession.

## 6 | DATA AND SCRIPT ACCESSIBILITY

[doi.org/10.13140/RG.2.2.18862.46407](https://doi.org/10.13140/RG.2.2.18862.46407)

## ACKNOWLEDGEMENTS

The authors would like to thank the Piaroa's Indigenous Community who allowed and contributed to this research in its traditional area. We also thank the National Institute of Agricultural Research (INIA-Amazonas) for research and logistics support. This research was funded by the Global Environment Facility (GEF-grant VEN/SGP/2010-2015); implementing agent United Nations Development Program, and the National Science and Technology Fund (FONACIT project number 2011000540). The first author received scholarships from Organization of American States (OAS) and PDJ from the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), Brazil.

## REFERENCES

APG IV, (2016). An update of the Angiosperm Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 141, 399–436.

Arroyo-Kalin, M., (2012). Slash-burn-and-churn: Landscape history and crop cultivation in pre-Columbian Amazonia. *Quatern. Int.* 249, 4–18.

Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Mac Nally, R., Thomson, J.R., Ferraz, S.F.B., Louzada, J., Oliveira, V.H.F., et al., (2016). Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* 535, 144–147.

Colwell, R.K., Chao, A., Gotelli, N.J., Lin, S.Y., Mao, C.X., Chazdon, R.L., Longino, J.T., (2012). Models and estimators linking individual-based and sample based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* 5, 3–21.

Chao, A., Jost, L., (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology* 93, 2533–2547.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* 84, 45–67.

Chazdon, R.L., (2014). *Second growth: The promise of tropical forest regeneration in an age of deforestation.* University of Chicago Press, Chicago, Illinois, USA.

Finegan, B., (1996). *Pattern and process in neotropical secondary rain forests: The first 100 years of*

succession. *Trends Ecol. Evol.* 11, 119-124.

Hadley, W., (2015). R ggplot2 package: An implementation of the grammar of graphics. <http://ggplot2.org>, <https://github.com/hadley/ggplot2>.

Hsieh, T.C., Ma, K.H., Chao, A., (2016). iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.12. <https://cran.r-project.org/web/packages/iNEXT/iNEXT.pdf>.

Jakovac, C.C., Peña-Claros, M., Kuyper, T.W., Bongers, F., (2015). Loss of secondary-forest resilience by land-use intensification in the Amazon. *J. Ecol.* 103, 67–77.

Jost, L., (2007). Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439.

Poorter, L., Rozendaal D. M., Bongers F., Almeida-Cortez J. S. et al. (2019). Wet and dry tropical forests show opposite successional pathways in wood density but converge over time. *Nature Ecology Evolution*, 3: 928-934.

R Core Team, (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rozendaal, D. M. A., Bongers, F., Aide, T. M., Alvarez-Dávila, E., Ascarrunz, N., Balvanera, P., Becknell, J. M., Bentos, T. V., Brancalion, P. H. S, Cabral, G. A. L., Calvo-Rodriguez, S., Chave, J., et al. (2019). Biodiversity recovery of Neotropical secondary forests. *Science Advance*, 5:3114.

Villa, P.M., Martins, S.V., Oliveira Neto, S.N., Rodrigues, A.C., Martorano, L., Delgado, L., Cancio, N.M., Gastauerg, M., (2018a). Intensification of shifting cultivation reduces forest resilience in the northern Amazon. *For. Ecol. Manage.* 430, 312-320.

Villa, P.M., Martins, S.V., Oliveira Neto, S.N., Rodrigues, A.C., Vieira, N., Delgado, L., Cancio, N.M., (2018b). Woody species diversity as an indicator of the forest recovery after shifting cultivation disturbance in the northern Amazon. *Ecol. Indic.* 95, 687-694.

Villa, P.M., Martins, S.V., Oliveira Neto, S.N., Rodrigues, A.C., (2017). Anthropogenic and biophysical predictors of deforestation in the Amazon: towards integrating REDD+ activities. *Bosque* 38, 433-446.

Zhang, K., de Almeida Castanho, A.D., Galbraith, D.R. et al. (2015). The Fate of Amazonian Ecosystems over the Coming Century Arising from Changes in Climate, Atmospheric CO<sub>2</sub>, and Land Use. *Glob. Chang. Biol.* 21, 2569–2587.

## PHYTOSOCIOLOGY IN R: A ROUTINE TO ESTIMATE PHYTOSOCIOLOGICAL PARAMETERS

Data de aceite: 12/02/2020

### Gustavo Heringer

Programa de Pós-Graduação em Ecologia Aplicada, Universidade Federal de Lavras - UFLA, Lavras, Minas Gerais, Brazil.

Laboratório de Ecologia e Evolução de Plantas - LEEP, Programa de Pós-graduação em Botânica, Departamento de Biologia Vegetal, Universidade Federal de Viçosa - UFV, Viçosa, Minas Gerais, Brazil.

### Pedro Manuel Villa

Laboratório de Restauração Florestal - LARF, Departamento de Engenharia Florestal, Universidade Federal de Viçosa - UFV, Viçosa, Minas Gerais, Brazil.

### Andreza Viana Neri

Laboratório de Ecologia e Evolução de Plantas - LEEP, Programa de Pós-graduação em Botânica, Departamento de Biologia Vegetal, Universidade Federal de Viçosa - UFV, Viçosa, Minas Gerais, Brazil.

**ABSTRACT:** The estimate of phytosociological parameters is one of the most useful approaches in the study of plant communities. There are several options for running a phytosociological analysis, such as the software FITOPAC, TURBOVEG, JUICE, and Mata Nativa, or another alternative is entering equations into a spreadsheet editor. Nevertheless, we consider

the recent rise in the number of users of the R program in the ecology field and, therefore, present here a code to run phytosociological parameters in R from both woody or alternatively for Braun-Blanquet's sample method. The code was created to allow the user, mainly beginners in the language, to upload the field data and analyze the phytosociological parameters as simply and directly as possible.

**KEYWORDS :** Abundance, vegetation cover, vegetation structure, importance value index, dominance

### FITOSSOCIOLOGIA NO R: UMA ROTINA PARA ESTIMAR PARÂMETROS FITOSSOCIOLOGICOS

**RESUMO:** A estimativa de parâmetros fitossociológicos estão entre as mais úteis abordagens no estudo de comunidades vegetais. Existem, portanto, algumas alternativas para rodar análises fitossociológicas, tais como os *softwares* FITOPAC, TURBOVEG, JUICE e Mata Nativa, ou inserir as equações em um editor de tabelas. No entanto, nós levamos em conta o aumento no número de usuários do programa R na área de ecologia e, portanto, apresentamos aqui um código para rodar parâmetros fitossociológicos no R para os

métodos de amostragem de espécies lenhosas ou alternativamente utilizando o método de Braun-Blanquet. O código em R foi criado para permitir aos usuários, em especial os iniciantes na linguagem, abrir os dados obtidos em campo e analisar parâmetros fitossociológicos de forma mais simples e direta possível.

**PALAVRA-CHAVE:** Abundância, cobertura vegetal, estrutura da vegetação, índice de valor de importância, dominância

## 1 | INTRODUCTION

Phytosociological parameters are widely applied in ecology with the aim of characterizing vegetation and understanding processes in plant communities. Nowadays, phytosociology is not limited to vegetation description, but it is the base for understanding patterns of community assemblage. Hence, it is very important to comprehend the structure of a plant community that is being investigated in order to delve further into questions concerning plant ecology. Phytosociology is a science that present results of a current assemblage, based on floristic composition. However, vegetation is not static and changes in space and time as a result of different ecological processes. For that reason, it is necessary to continually update phytosociological studies (CÁCERES et al., 2015).

Phytosociological studies started with the botanist Josias Braun-Blanquet around the 1920s wherein plant community sampling in plots was standardized and based on the information he used to name the vegetation (DENGLER, 2017). Nowadays, it is a method that widely used in Europe, northern Asia, Africa and Latin America. This method is over 100 years old and data have been accumulating in published and unpublished studies done on several million research plots; however, some of those data have been registered in compound vegetation-plots databases (DENGLER et al., 2011).

Phytosociological parameters describe the frequency, abundance, and cover of each species in relation to the community (BRAUN-BLANQUET, 1979; MORO & MARTINS, 2013; MUELLER-DOMBOIS & ELLENBERG, 1974; MUNHOZ & ARAÚJO, 2013), and are divided into absolute and relative types. The absolute parameters are related to the total area sampled or total number of samplings, while the relative ones concern the total frequency, total abundance, and total basal area (BRAUN-BLANQUET, 1979; MUELLER-DOMBOIS; ELLENBERG, 1974). As we can see by the names, these parameters are quite intuitive and very informative regarding vegetation structure.

Besides describing the vegetation, many studies use phytosociological parameters to investigate patterns of distribution of a focal species between samples (KOUR & SHARMA, 2014; WHITFORD, 1949), the relationship of dominance within a

vegetation community (JIBRIN et al., 2018; RODRIGUES; VILLA; NERI, 2019), or to investigate the set of species that have greater weight in the plant community structure, for ecological restoration, or for human use (MELI et al., 2014; SOLDATI et al., 2017). Sometimes these parameters are used to implement additional metrics to investigate the vegetation community, such as the application of relative abundance and relative frequency in the measurement of indicator species, as proposed by DUFRÊNE and LEGENDRE (1997). In addition to being used as part of many scientific studies in plant ecology, at least in some countries, phytosociological parameters are required in vegetation studies that are required as part of the process of environmental licensing and in the preparation of the management plan of protected areas.

Because of the relevance of the phytosociological studies, there is software that can process these calculations, such as FITOPAC, TURBOVEG, JUICE, and Mata Nativa (respectively, SHEPHERD, 1996; HENNEKENS & SCHAMINÉE, 2001; TICHÝ, 2002; CIENTEC, 2016), and these benefit the researcher because the use of a spreadsheet editor and manual insertion of equations is not necessary. Although some software is freely available, the usage of different platforms during research can cause inconvenience. For instance, often times a researcher must become familiar with different templates, deal with distinct input tables, and export the data to another software type if he wants to proceed with more complex analyses. Hence, development of a code in the software R will favor researchers that already use R as their main statistical software, reducing time wasted in dealing with other platforms or exporting data. Moreover, a simple code will also facilitate the beginning user in the interpretation of the parameters, as well as in the understanding of the R language. Undoubtedly, the R program is the most used software in ecological studies today and many researchers will benefit from running phytosociological parameters and statistical analyses in the same environment.

In this context, we aimed to estimate phytosociological parameters of plant species communities using the R software. Here we present a set of command lines wrote in R that can be applied for calculation of phytosociological parameters. Our intent was to create a code that allows a researcher to 1) update the field table, 2) run the parameters, 3) and put the table with all phytosociological parameters into an appropriate format.

## 2 | METHODS

To implement phytosociological analysis, we consider the most common types of field tables used for woody species sampling and Braun-Blanquet cover-abundance sampling. The formats of both tables are presented in their own code (see below).



## 2.1 Woody sampling

For woody species sampling, our code was built to transform circumference at breast height (cbh) into basal area, and subsequently calculate the sum of basal area, frequency and abundance per species. These values will be used with the total area sampled (in hectares) and the number of areas sampled in the following equations.

Absolute dominance:

$$ADo = \frac{BA_i}{\text{area (ha)}} \quad \text{Eq. 1}$$

Where,  $BA_i$  is the basal area of the species  $i$ , and area is the total area sampled in hectare.

Relative dominance:

$$RDo = 100 \left( \frac{BA_i}{\sum_i^n BA_i} \right) \quad \text{Eq. 2}$$

Where,  $BA_i$  is the basal area of the species  $i$ .

Absolute frequency:

$$AFr = \frac{Fr_i}{Nid} \quad \text{Eq. 3}$$

Where,  $Fr_i$  is the number of sample areas where the species  $i$  occurs and  $Nid$  is the number of sampling areas.

Relative frequency:

$$AFr = 100 \left( \frac{Fr_i}{\sum_i^n Fr_i} \right) \quad \text{Eq. 4}$$

Where,  $Fr_i$  is the number of sample areas where the species  $i$  occurs.

Absolute density:

$$ADe = \frac{n_i}{\text{area (ha)}} \quad \text{Eq. 5}$$

Where,  $n_i$  is the number of individuals of each species and area is the total area sampled in hectares.

Relative density:

$$RDe = 100 \left( \frac{n_i}{N} \right) \quad \text{Eq. 6}$$

Where,  $n_i$  is the number of individuals of each species and  $N$  is the total abundance.

Importance value:

$$IV = RDo + RFr + RDe \quad \text{Eq. 7}$$

## 2.2 Cover value and abundance scale

For the Braun-Blanquet approach, the code is divided into two main sets. Firstly, to calculate proportion of cover, frequency and abundance per species that will later be applied in the equations to get the phytosociological parameters. The parameters for absolute and relative dominance are replaced with cover value and relative cover (calculated from proportional cover, instead of basal area) and used in the following equation with the total area sampled (in square meters) and the number of areas used to calculate cover value and relative cover.

Cover value:

$$CV = \frac{C_i}{\text{area (m}^2\text{)}} \quad \text{Eq. 8}$$

Where,  $C_i$  is the proportional cover of the species  $i$ .

Relative cover:

$$RC = 100. \left( \frac{C_i}{\sum_i^n C_i} \right) \quad \text{Eq. 9}$$

Where,  $C_i$  is the proportional cover of the species  $i$ .

Here, we are considering the Braun-Blanquet method used for small areas and square meters as a unit of measure. Therefore, calculations of the parameters for frequency and density are the same as presented in equations 3 to 6 but use the sampled area in square meters instead of hectares to calculate absolute density (Eq. 5). Importance value is calculated as presented in Eq. 7 but relative dominance (RDo) is replaced with relative cover (RC).

## 3 | RESULTS AND DISCUSSION

The code created to run phytosociological parameters analysis in R is available at the GitHub platform (<https://github.com/GustavoHeringer/phytosociology>) and is transcribed below in separated sections. Our intent here is to highlight the main steps used in the code and to provide a code that is as useful as possible.

### 3.3 Code for woody species sampling

Firstly, the researcher must prepare an input table like Table 1. The table presented here is largely used to record woody vegetation with one small adaptation, which is that the researcher must add a column to report the abundance of species. In woody species sampling, usually each line in the table refers to an individual and individuals with more than one trunk receive more than one value for the circumference at breast height (cbh) column. Here, we propose to use one line for each trunk and, therefore, an individual with more than one trunk will occupy as many lines as there

exist trunks. Hence, this individual must receive abundance as 1, in the recording of the first trunk, and receive 0 for the other trunks (see sp04 in Table 1). Thus, the calculation of basal area and abundance per individual will be performed correctly.

plot	species	abundance	cbh	height
81	sp01	1	17.9	4.0
81	sp02	1	29.8	5.0
81	sp03	1	84.6	10.8
81	sp04	1	40.6	7.5
81	sp04	0	21.8	3.9

Table 1: A common example of a data frame used in fieldwork for woody sampling that must be uploaded in R.

\*cbh = circumference at breast height

Another mandatory step when running the code is to insert the total area in hectares sampled and the number of sample areas, as shown below (0.1 ha and 10 sample areas).

```
area.wo <- 0.1
sample.wo <- 10
```

Subsequently, the user must run a set of codes to prepare a data frame that will be used in the output and to calculate the phytosociological parameters. At the same time, the circumference at breast height in centimeters is converted to basal area in meters squared. If the user sampled diameter at breast height (dbh) instead cbh, the first code line below must be replaced with “df.wo\$basal\_a <- 3.1415926\*(df.wo\$dbh/2)^2/10000”.

```
df.wo$basal_a <- (3.1415926*(df.wo$cbh/(2*3.1415926))^2)/10000
basal_a_wo <- tapply(df.wo$basal_a, df.wo$species, function(x) sum(x))
freq_wo <- tapply(df.wo$plot, df.wo$species, function(x) length(unique(x)))
abund_wo <- tapply(df.wo$abundance, df.wo$species, function(x) sum(x))
table_wo <- data.frame(SPECIES=names(abund_wo), BASAL_A=basal_a_wo,
FREQ=freq_wo, ABUND=abund_wo)
```

In the following set of codes, the phytosociological parameters will be calculated and included in the data frame.

```
table_wo$ADo <- table_wo$BASAL_A/area.wo
table_wo$RDo <- 100*(table_wo$BASAL_A/sum(table_wo$BASAL_A))
table_wo$AFr <- 100*(table_wo$FREQ/sample.wo)
table_wo$RFr <- 100*(table_wo$FREQ/sum(table_wo$FREQ))
table_wo$ADe <- table_wo$ABUND/area.wo
table_wo$RDe <- 100*(table_wo$ABUND/sum(table_wo$ABUND))
```

```
table_wo$IVI <- table_wo$RDo + table_wo$RFr + table_wo$RDe
```

Finally, we provide a very simple code to export the output as a .txt format that can be used in any other software.

```
write.table(table_wo, "phyto_from_field_to_table_woody.txt", row.names = F)
```

### 3.4 Code for cover-abundance sampling

Similar to woody species sampling, the user must insert input data in a standardized format (Table 2). The table that we propose is commonly applied in the Braun-Blanquet approach. Frequently, the researchers that use the Braun-Blanquet cover-abundance scale approach do not record the abundance of each species; this occurs mainly in vegetation communities composed by grass where it can be very difficult to distinguish individuals. Even so, we also provided a code that allows for the calculation of absolute and relative densities and the creation of these variables. Thus, researchers that need to calculate the phytosociological parameters with respect to density must account for abundance during fieldwork.

plot	species	abundance	cover class	cover percentage	cover proportion
81	sp01	4	2	15.0	0.150
81	sp02	1	R	00.1	0.001
81	sp03	2	2	15.0	0.150
81	sp04	8	5	87.5	0.875

Table 2: A common example of a data frame used in fieldwork for cover-abundance sampling that must be uploaded in R.

\*cover class = cover class *sensu* BRAUN-BLANQUET (1979); cover percentage = mean of cover in percentage based on cover class.

Again, the user must create the two objects regarding to the area sampled, here in square meters and the number of sample areas.

```
area.bb <- 10  
sample.bb <- 10
```

Subsequently, there are three sets of codes. The first one, to prepare the data frame, the second to calculate the phytosociological parameters, and, finally, the code to export the output in a .txt format.

```
cover_prop <- tapply(df.bb$cover_prop, df.bb$species, function(x) sum(x))  
freq <- tapply(df.bb$plot, df.bb$species, function(x) length(unique(x)))  
abund <- tapply(df.bb$abundance, df.bb$species, function(x) sum(x))  
table_bb <- data.frame(SPECIES=names(abund), COVER=cover_prop, FREQ=freq,  
ABUND=abund)
```

```

table_bb$CV <- 100*(table_bb$COVER/area.bb)
table_bb$RC <- 100*(table_bb$COVER/sum(table_bb$COVER))
table_bb$AFr <- 100*(table_bb$FREQ/sample.bb)
table_bb$RFr <- 100*(table_bb$FREQ/sum(table_bb$FREQ))
table_bb$ADe <- table_bb$ABUND/area.bb
table_bb$RDe <- 100*(table_bb$ABUND/sum(table_bb$ABUND))
table_bb$IVI <- table_bb$RC + table_bb$RFr + table_bb$RDe

```

```
write.table(table_bb, "phyto_from_field_to_table_braun_b.txt", row.names = F)
```

### 3.5 Output format

For woody samples, the output table will contain species name, abundance, basal area (in meters squared), frequency, and the phytosociological parameters including importance value index per hectare (Table 3).

SPECIES	BASAL_A	FREQ	ABUND	ADo	RDo	AFr	RFr	ADe	RDe	IVI
sp01	0.11	7	16	1.10	2.54	70	4.17	160	3.33	10.04
sp02	0.23	10	46	2.32	5.38	100	5.95	460	9.58	20.92
sp03	0.44	9	20	4.41	10.22	90	5.36	200	4.17	19.75
sp04	0.04	3	5	0.43	0.99	30	1.79	50	1.04	3.82

Table 3: An example of output of the code in .txt file extension. BASAL\_A = Basal area, FREQ = Frequency, ABUND = Abundance, ADo = Absolute Dominance, RDo = Relative Dominance, AFr = Absolute Frequency, RFr = Relative Frequency, ADe = Absolute Density, RDe = Relative Density, IVI = Importance Value Index.

For Braun-Blanquet cover-abundance methods the output will contain almost the same result as for woody samples. Nevertheless, it replaces basal area, absolute dominance, and relative dominance with cover, cover value, and relative cover, respectively (Table 4). Furthermore, in this case all parameters will be reported per meter squared instead of hectares.

SPECIES	COVER	FREQ	ABUND	CV	RC	AFr	RFr	ADe	RDe	IVI
sp01	0.15	2	5	1.51	2.03	20	3.33	0.5	5.21	10.57
sp02	0.33	5	13	3.32	4.46	50	8.33	1.3	13.54	26.33
sp03	1.58	5	8	15.80	21.22	50	8.33	0.8	8.33	37.88
sp04	0.39	5	5	3.90	5.24	50	8.33	0.5	5.21	18.78

Table 4: An example of output of the code in .txt file extension. FREQ = Frequency, ABUND = Abundance, CV = Cover Value, RC = Relative Cover, AFr = Absolute Frequency, RFr = Relative Frequency, ADe = Absolute Density, RDe = Relative Density, IVI = Importance Value Index.

Both .txt tables can be easily opened in any software or spreadsheet editor and



the researcher can use this code for many different objectives, e.g., to use the output of phytosociology as a variable in future analyses. The researcher can also maintain their study in R as the objects “table\_wo” and “table\_bb” for later analysis.

As a final highlight, the researcher may decide to use a different heading in the field table, for instance “ProportionalCover” instead of “cover\_prop” as provided in our code. To this end there are two equally easy options. First, the researcher can just change the heading in the original field table or in R and run the code normally. As a second alternative, the term “df.bb\$cover\_prop” can be substituted for “df.bb\$ProportionalCover”, throughout the entire code.

## 4 | FINAL CONSIDERATIONS

We presented a set of command lines that allow to calculate phytosociological parameters exclusively in R. Therefore, R users can apply phytosociological approach in their studies without move from other planforms. We also highlight that our code provides more flexibility to the researchers, since allows the user to calculate phytosociological parameters from woody and Braun-Blanquet approach (cover sampling). Finally, we believe this routine can be easily applied by beginning users and will favor a better understanding of phytosociological parameters, while, at the same time, can work as a gateway to the R language.

## REFERENCES

BRAUN-BLANQUET, Josias. **Fitosociología: Bases para el estudio de las comunidades vegetales**. Madrid: H. Blume, 1979.

CÁCERES, Miquel et al. A comparative framework for broad-scale plot-based vegetation classification. **Applied Vegetation Science**, [s. l.], v. 18, n. 4, p. 543–560, 2015. Disponível em: <<http://doi.wiley.com/10.1111/avsc.12179>>. Acesso em: 8 dez. 2019.

CIENTEC - Consulting and Development Systems. Software Mata Nativa 4. Viçosa: CIENTEC, 2016.

DENGLER, Jürgen et al. The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. **Journal of Vegetation Science**, [s. l.], v. 22, n. 4, p. 582–597, 2011. Disponível em: <<http://doi.wiley.com/10.1111/j.1654-1103.2011.01265.x>>. Acesso em: 8 dez. 2019.

DENGLER, Jürgen. Phytosociology. In: RICHARDSON, Douglas et al. (Eds.). *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Oxford: John Wiley & Sons, 2017. p. 1–6.

DUFRENE, Marc; LEGENDRE, Pierre. DufreneLegendre1997\_species assemblages and indicator species\_the need for flexible asymmetrical approach.PDF. **Ecological Monographs**, [s. l.], v. 67, n. 3, p. 345–366, 1997.

HENNEKENS, Stephan M.; SCHAMINÉE, Joop H. J. TURBOVEG, a comprehensive data base management system for vegetation data. **Journal of Vegetation Science**, [s. l.], v. 12, n. 4, p. 589–591, 2001.

JIBRIN, A. et al. Phytosociological analysis of woody plant species as determinant of above ground carbon stock in the Guinea Savanna Ecological Zone of Nigeria. **FUTY Journal of the Environment**, [s. l.], v. 12, n. 2, p. 56–65, 2018.

KOUR, Kulvinder; SHARMA, Sanjay. Diversity and Phytosociological Analysis of Tree Species in Sacred Groves of Vijaypur Block , Samba ( J & K ). [s. l.], v. 3, n. 6, p. 2012–2015, 2014.

MELI, P. et al. Combining ecological, social and technical criteria to select species for forest restoration. **Applied Vegetation Science**, [s. l.], v. 17, n. 4, p. 744–753, 2014.

MORO, Marcelo Freire; MARTINS, Fernando Roberto. Métodos de levantamento do componente arbóreo-arbustivo. In: FELFILI, Jeanine Maria et al. (Eds.). **Fitossociologia no Brasil: métodos e estudos de casos. Volume I**. Viçosa: Editora UFV, 2013. p. 174–212.

MUELLER-DOMBOIS, Dieter; ELLENBERG, Heinz. **Aims and methods of vegetation ecology**. New York: John Wiley & Sons, 1974.

MUNHOZ, C. B. R.; ARAÚJO, G. M. Métodos de amostragem do estrato herbáceo-subarbusivo. In: FELFILI, Jeanine Maria et al. (Eds.). **Fitossociologia no Brasil: métodos e estudos de casos. Volume I**. Viçosa: Editora UFV, 2013. p. 213–230.

RODRIGUES, Alice Cristina; VILLA, Pedro Manuel; NERI, Andreza Viana. Fine-scale topography shape richness, community composition, stem and biomass hyperdominant species in Brazilian Atlantic forest. **Ecological Indicators**, [s. l.], v. 102, n. May 2018, p. 208–217, 2019. Disponível em: <<https://doi.org/10.1016/j.ecolind.2019.02.033>>

SHEPHERD, G.J. **FITOPAC versão 2.0**. Campinas: UNICAMP, 1996.

SOLDATI, Gustavo Taboada et al. How do people select plants for use? Matching the Ecological Apparency Hypothesis with Optimal Foraging Theory. **Environment, Development and Sustainability**, [s. l.], v. 19, n. 6, p. 2143–2161, 2017.

TICHÝ, Lubomír. JUICE, software for vegetation classification. **Journal of Vegetation Science**, [s. l.], v. 13, n. 3, p. 451–453, 2002.

WHITFORD, Philip B. Distribution of Woodland Plants in Relation to Succession and Clonal Growth. **Ecology**, [s. l.], v. 30, n. 2, p. 199–208, 1949.

## ANÁLISE DE DADOS DE DESMATAMENTO COM R: VISUALIZAÇÃO INTERATIVA COM SHINY

Data de aceite: 12/02/2020

### Carlos Eduardo Cardoso

Academia da Força Aérea  
Pirassununga-SP

### Mauricio Evandro Eloy

Academia da Força Aérea  
Pirassununga-SP

### João Paulo Martins dos Santos

Academia da Força Aérea  
Pirassununga-SP

### Alessandro Firmiano de Jesus

Academia da Força Aérea  
Pirassununga-SP

**RESUMO:** O uso do software R tem crescido nos últimos anos, especialmente em trabalhos envolvendo processamento de dados e análises estatísticas. Seu dinamismo e robustez permitem uma análise detalhada de diferentes tipos de situações e, em muitos casos, a interação entre o desenvolvedor do projeto e o usuário final. Neste trabalho serão apresentadas algumas ferramentas para a realização de análises estatísticas usando o R com dados sobre o desmatamento da Amazônia Legal. Essa base apresenta as áreas de desmatamento, em km<sup>2</sup>, no período de 2004 a 2019. Os resultados estão concentrados na análise exploratória de dados e na visualização interativa por meio de

aplicativos.

**PALAVRAS-CHAVE:** R. Análise Exploratória. Shiny. Desmatamento.

### DEFORESTATION DATA ANALYSIS WITH R: SHINY INTERACTIVE VIEWING

**ABSTRACT:** The use of R software has increased in recent years, especially in works involving data processing and statistical analysis. Its dynamism and robustness allow a detailed analysis of different types of situations and, in many cases, the interaction between the project developer and the end user. In this paper we will present some tools for performing statistical analyzes using R with data on deforestation of the Legal Amazon. This database presents the deforestation areas, in km<sup>2</sup>, from 2004 to 2019. The results are concentrated in exploratory data analysis and interactive visualization through applications.

**KEYWORDS:** R. Exploratory Analysis. Shiny. Deforestation.

## 1 | INTRODUÇÃO

O uso do software R vem crescendo entre profissionais da área de estatística, cientistas de dados e pesquisadores ao longo dos últimos anos, conforme evidenciado pelos dados da

figura 1. Essa evidência é corroborada pela disponibilidade de 15.286 pacotes (R Core Team, 2018). Detalhes adicionais sobre a disponibilidade de pacotes com uma extensa lista contemplando diversas áreas do conhecimento podem ser encontrados em (R Development Core Team, 2019).

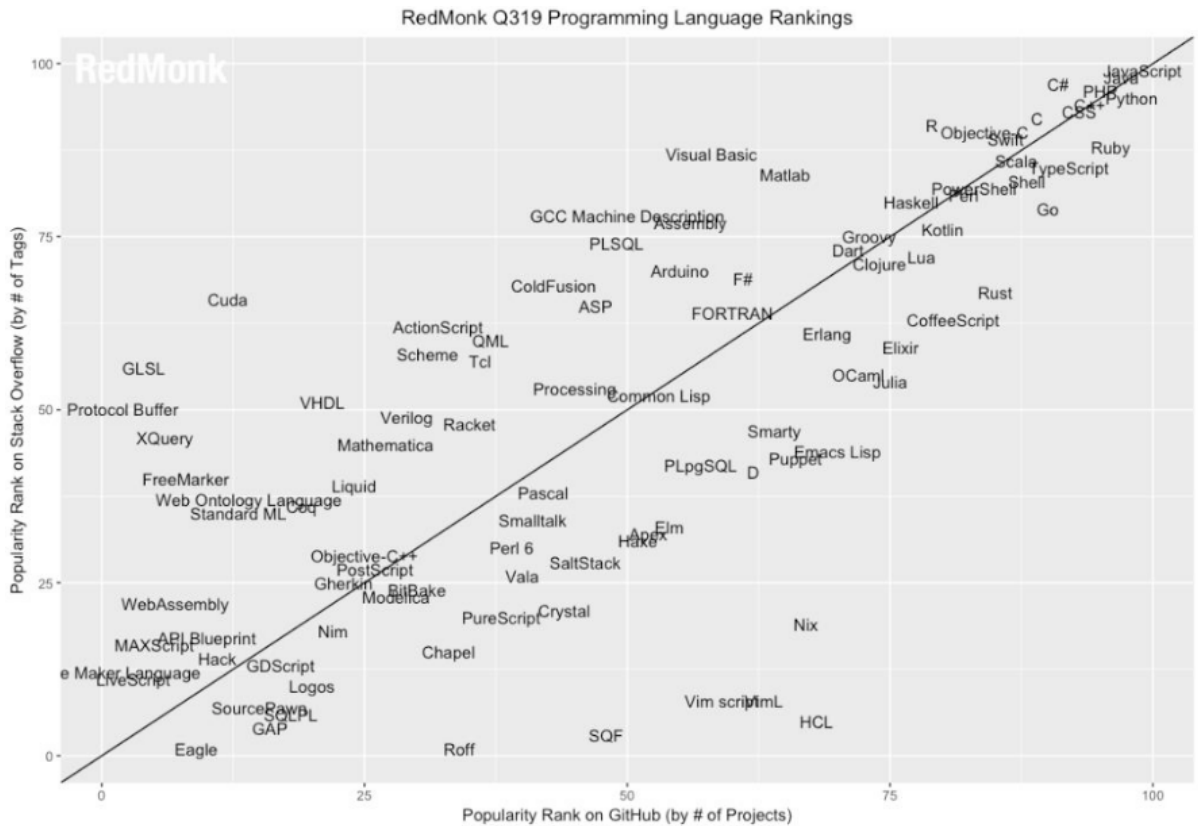


Figura 1: Uso do R como linguagem de programação. Fonte: The RedMonk Programming Language Rankings.

A disponibilidade de pacotes associada à facilidade de instalação, auxilia na expansão do uso do software R tanto no ambiente acadêmico como no ambiente corporativo. Estatística descritiva, teste de hipóteses, modelos de regressão (linear e não linear) e ANOVA, por exemplo, juntamente com a capacidade de visualização de resultados do ggplot2 (WICKHAM, 2018) são algumas tarefas que podem ser executadas no ambiente R.

A interatividade no R pode ser adicionada por meio do uso do pacote Shiny (CHANG et al., 2019), o qual possibilita a construção de interfaces, que serão denominadas aplicativos (*apps*) interativos.

O objetivo é apresentar alguns pacotes para uma análise exploratória dos dados, utilizar mapas de desmatamento ao longo dos últimos anos na região da Amazônia Legal e construir *apps* interativos para descrever, analisar os dados coletados e fazer previsões de áreas de desmatamento para os próximos anos. Os dados públicos analisados fazem referência ao desmatamento nessa área no período de 2004 a 2019.

A análise estatística básica combinada com a apresentação gráfica dinâmica

possibilita uma visão mais ampla para caracterizar o fenômeno do desmatamento. Dentre os impactos causados pelo desmatamento destacam-se perdas de serviços e bens ecossistêmicos e até mesmo mudanças climáticas como, por exemplo, o aumento da frequência do fenômeno EL Niño (FEARNSIDE, 2006). Os resultados numéricos e as representações gráficas, incluindo os *apps*, mostram a evolução das áreas desmatadas. A visualização gráfica interativa permite traçar o perfil da evolução desse problema. Dessa forma, proporciona a construção mais coerente de argumentos para a descrição do fenômeno e, por consequência, permite inferir o tratamento dado à questão principalmente quanto à conscientização social e às políticas públicas de combate ao desmatamento. A seguir, será apresentada uma breve descrição dos pacotes utilizados bem como os resultados numéricos e gráficos.

## 2 | OS PACOTES GGPLOT2, LEAFLET, FORECAST E SHINY

Conforme ilustrado em (R Development Core Team, 2019), o software R trabalha com diferentes tipos de pacotes que podem ser instalados e carregados de acordo com a necessidade do usuário. Alguns desses pacotes serão aqui destacados em razão da sua eficiência e aplicabilidade.

### 2.1 GGPLOT2

O *ggplot2* é o principal pacote gráfico do R. É um dos membros centrais do pacote *tidyverse* (WICKHAM, 2017). Com o *ggplot2* é possível criar gráficos mais elegantes e versáteis (WICKHAM, 2018).

### 2.2 LEAFLET

O pacote *Leaflet* (CHENG et al., 2018) permite a criação de mapas interativos diretamente no R. Trata-se de uma biblioteca JavaScript que produz gráficos com diferentes interfaces e permite a interação com o usuário por meio das funções que carrega.

### 2.3 FORECAST

No pacote *Forecast* (HYNDMAN, et al., 2019) um modelo ARIMA para séries temporais pode ser obtido por meio da busca automática dos parâmetros. Este pode ser utilizado para realizar previsões e evidenciar o processo de desmatamento para os próximos anos.

### 2.4 SHINY

Finalmente, o pacote *shiny* (CHANG et al., 2019) permite a criação dos



chamados *apps*: aplicativos interativos em R para navegadores (SCHMULLER, 2018). Esses *apps* tornam possível o compartilhamento de dados e análises que podem ser prontamente executados. Vale ressaltar que a utilização de um *app* shiny está associado à execução do software R.

### 3 | ANÁLISE EXPLORATÓRIA DOS DADOS

Os dados coletados para a realização dessas análises são provenientes do projeto PRODES que monitora as áreas de desmatamento na Amazônia Legal através de imagens de satélite. Entende-se como Amazônia Legal uma área, de aproximadamente, 5.217.423 km<sup>2</sup> e que engloba os estados do Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins e Maranhão. Os dados utilizados estão disponíveis no site:

<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>.

Com base nesses dados coletados via satélite, o governo brasileiro estabelece políticas públicas para combater o desmatamento nessa região.



Figura 2: Amazônia Legal usando o pacote LeafLet

A Figura 2 ilustra um exemplo de mapa gerado com o pacote Leaflet e destaca os estados com compreendem a Amazônia Legal. É possível efetuar a visualização de níveis detalhados, ou seja, uma visualização tanto ao nível do país (ex: Brasil), quanto dos níveis de estado, cidade, bairro, etc. É possível também a criação de *popups*, que são pequenas janelas contendo informações sobre cada localidade, desde que estejam na base de dados.

### 3.1 Análise descritiva

Segundo Schmuller (2019, p.9) a estatística usa os números para a tomada de decisões. Para tanto, descrever os dados é uma parte importante desse processo. Essa análise inicial permite olhar para um grande conjunto de dados de forma mais específica, analisando seu comportamento e identificando tendências. Com o comando *summary* no R, é possível realizar uma estatística básica dos dados coletados. A base de dados é apresentada na Tabela 1 e os resultados obtidos a partir do comando *summary*, são apresentados na Tabela 2:

UF	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
ACRE	728	592	398	184	154	167	259	280	305	221	309	264	372	257	444	688
AMAZONAS	1232	775	788	610	604	405	595	502	523	583	500	712	1129	1001	1045	1421
AMAPÁ	46	33	30	39	100	70	53	66	27	23	31	25	17	24	24	8
MARANHÃO	755	922	674	631	1271	828	712	396	269	403	257	209	258	265	253	215
MATO GROSSO	11814	7143	4333	2678	3258	1049	870	1120	757	1139	1075	1601	1489	1561	1490	1685
PARÁ	8870	5899	5659	5526	5607	4281	3770	3008	1741	2346	1887	2153	2992	2433	2744	3862
RONDÔNIA	3858	3244	2049	1611	1136	482	435	865	773	932	684	1030	1376	1243	1316	1245
RORAIMA	311	133	231	309	574	121	256	141	124	170	219	156	202	132	195	617
TOCANTINS	158	271	124	63	107	61	49	40	52	74	50	57	58	31	25	21

Tabela 1: Dataframe com as áreas de desmatamento.

AC	AM	AP	MA	MT
Min. :167.0	Min. : 405.0	Min. : 8.00	Min. : 209.0	Min. : 757
1st Qu.:256.2	1st Qu.: 568.0	1st Qu.: 24.00	1st Qu.: 257.8	1st Qu.: 1109
Median :292.5	Median : 661.0	Median : 30.50	Median : 399.5	Median : 1526
Mean :357.6	Mean : 776.6	Mean : 38.50	Mean : 519.9	Mean : 2692
3rd Qu.:409.5	3rd Qu.:1012.0	3rd Qu.: 47.75	3rd Qu.: 722.8	3rd Qu.: 2823
Max. :728.0	Max. :1421.0	Max. :100.00	Max. :1271.0	Max. :11814
PA	RO	RR	TO	AMZ LEGAL
Min. :1741	Min. : 435	Min. :121.0	Min. : 21.00	Min. : 4571
1st Qu.:2411	1st Qu.: 842	1st Qu.:139.0	1st Qu.: 46.75	1st Qu.: 6365
Median :3389	Median :1190	Median :198.5	Median : 57.50	Median : 7500
Mean :3924	Mean :1392	Mean :243.2	Mean : 77.56	Mean :10021
3rd Qu.:5546	3rd Qu.:1435	3rd Qu.:269.2	3rd Qu.: 82.25	3rd Qu.:11966
Max. :8870	Max. :3858	Max. :617.0	Max. :271.00	Max. :27772

Tabela 2: Uso do comando *summary* no R

O dataframe da Tabela 1 contém os Estados da Amazônia Legal e as respectivas áreas de desmatamento, em km<sup>2</sup>. O comando *summary* retorna o menor valor desse conjunto, o primeiro quartil, mediana, média, terceiro quartil e o valor máximo para cada estado.

Uma análise visual dos resultados permite identificar que a menor área desmatada em um ano está no Amapá (8 km<sup>2</sup>) e a maior área desmatada em um ano está no Mato Grosso (11.814 km<sup>2</sup>); a maior média de desmatamentos pertence ao Pará (3.924 km<sup>2</sup>) e a menor média é a do Amapá (38,5 km<sup>2</sup>). Mato Grosso é o Estado que apresenta a segunda maior área de desmatamento, nesse período.

Um detalhe interessante do resumo descritivo da Tabela 2 é que 75% (3° Quartil) das áreas desmatadas em Mato Grosso correspondem a valores menores

que 2823 km<sup>2</sup>, que está bem distante da área máxima desmatada, que é de 11814 km<sup>2</sup>, registrada em 2004. Os valores registrados entre 2004 e 2008 estão muito acima dos demais registros. Ao calcularmos o desvio padrão (2940,2 km<sup>2</sup>) temos a comprovação dessa variabilidade. A Figura 3 nos mostra que Mato Grosso, Pará e Rondônia são os estados que apresentam maior variabilidade nas áreas desmatadas.

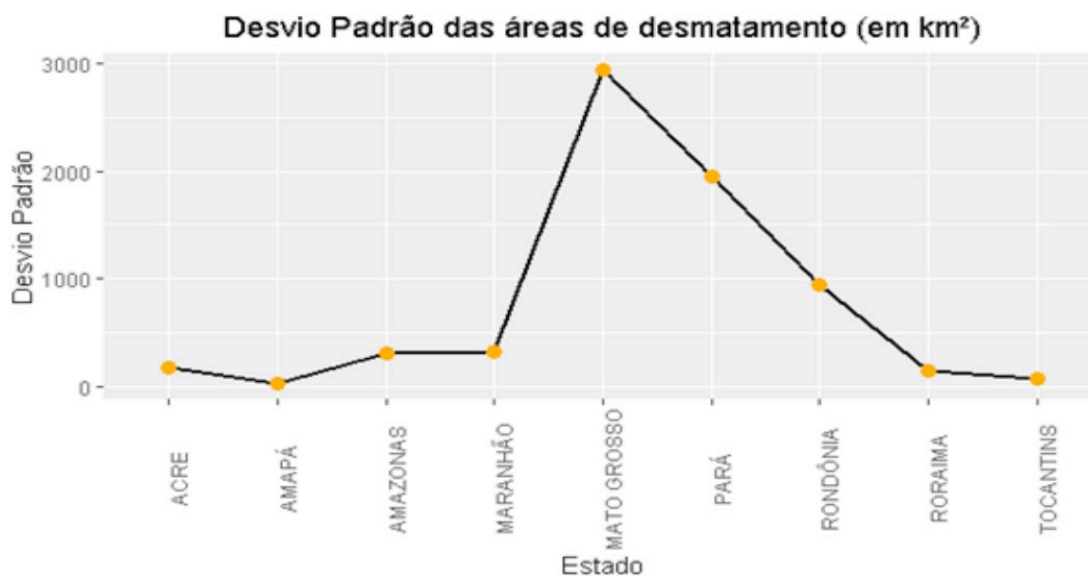


Figura 3: Desvio padrão

Uma comparação visual das áreas de desmatamento entre os estados é ilustrado pelo gráfico de barras, obtido por meio do ggplot2, mostrado na Figura 4. Amapá e Pará são os estados com menor e maior área desmatada, respectivamente. Um padrão similar poderia ser observado considerando-se as médias das áreas de desmatamento.

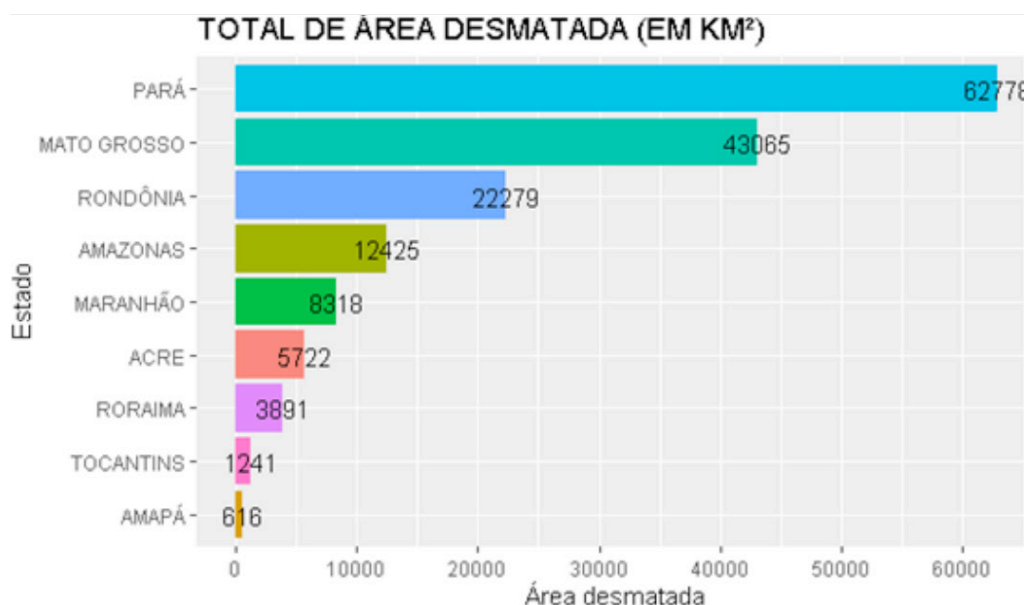


Figura 4: Área de desmatamento por estado usando o ggplot2.

Uma visão da evolução da área total desmatada no país pode ser obtida pela

análise da Figura 5, construída com o ggplot2. O ano de 2004 destaca-se como o ano com maior área desmatada no período observado. Em seguida, uma tendência decrescente no desmatamento pode ser observada até o ano de 2012. A partir disso, uma tendência ao crescimento. A reta  $y = 10.000 \text{ km}^2$  é a média das áreas desmatadas em todos os anos e em todos os Estados e serve como parâmetro para mostrar que a quantidade desmatada entre 2004 e 2008 é equivalente à quantidade desmatada nos demais anos.

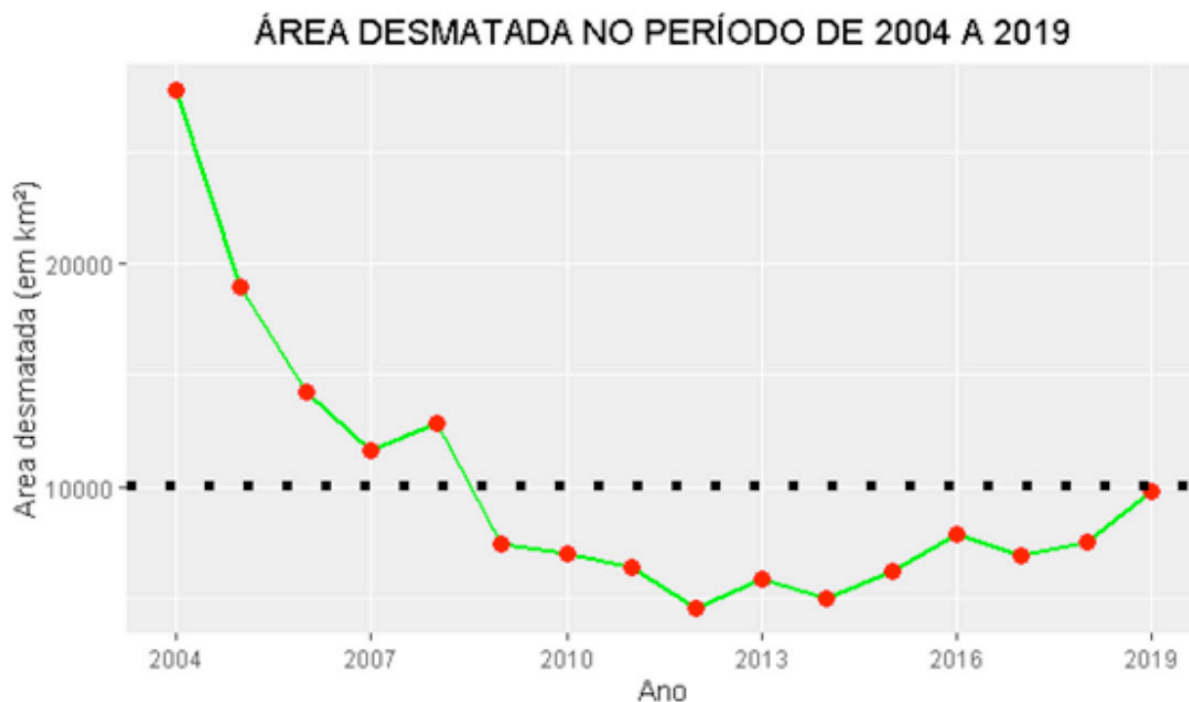


Figura 5: Evolução das áreas desmatadas

### 3.2 usando o leaflet

A utilização de mapas de calor é uma forma visual de apresentação e análise comparativa dos dados. Com o auxílio desses mapas é possível mostrar, de forma mais interativa, como se dá, ano a ano, o desmatamento da Amazônia Legal. A grande vantagem do Leaflet é a interação que o usuário pode realizar com a representação criada. A partir disso, diferentes informações podem ser extraídas da base de dados. Logo, o usuário pode navegar pelas informações e tirar suas próprias conclusões. Também é possível observar tendências e verificar padrões nos dados analisados.

A Figura 6 mostra os mapas de calor dos anos de 2004, 2005, 2006, 2010, 2014 e 2019. É possível criar um mapa para cada ano e realizar as comparações. É visível a redução na quantidade de área desmatada em Mato Grosso ao longo dos anos. Por outro lado, estados menores como o Acre, por exemplo, mantiveram valores muito próximos.

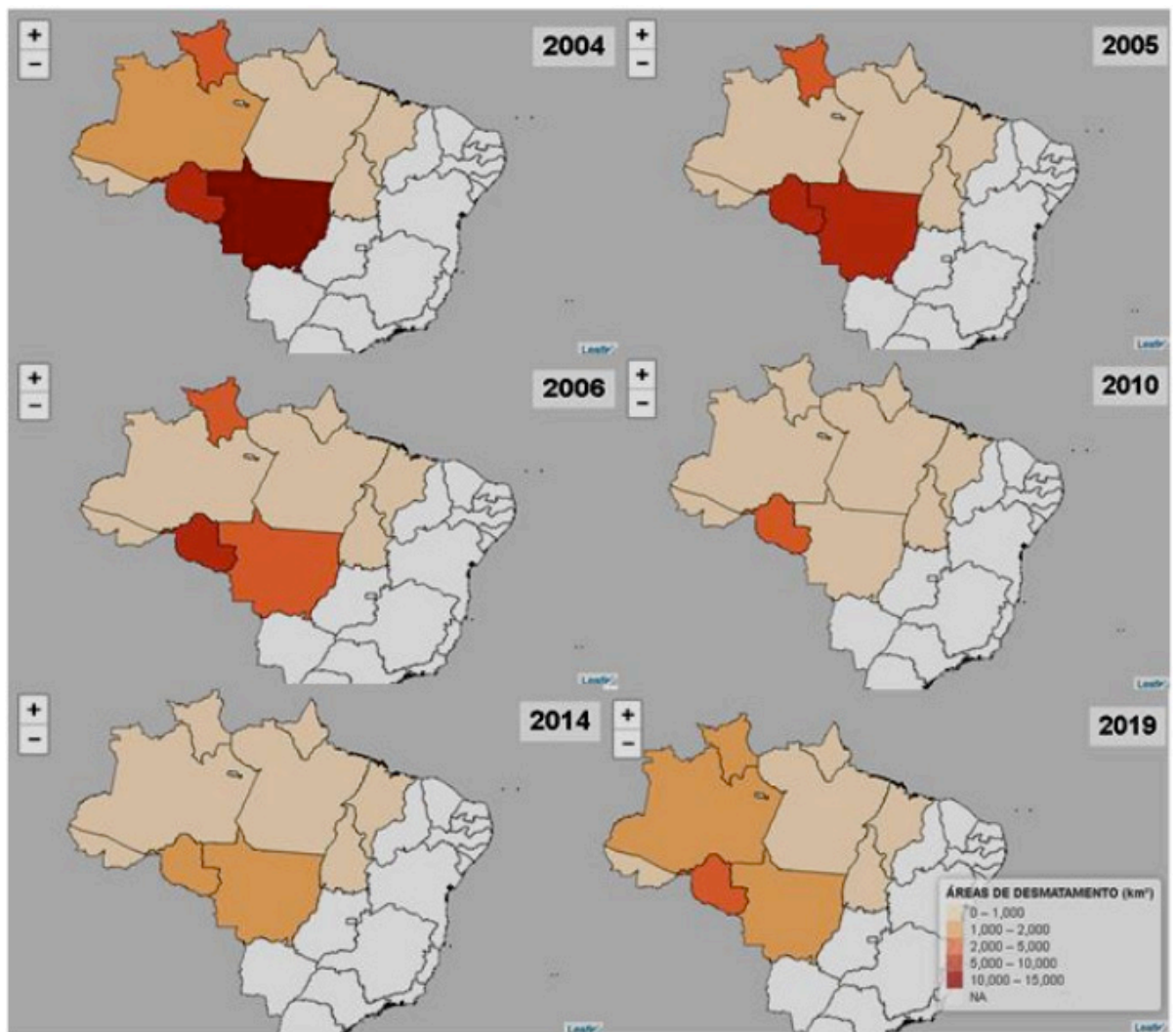


Figura 6: Comparativo do desmatamento

Esse tipo de representação permite analisar, através da escala de cores, os estados que possuem as maiores áreas de desmatamento, ano a ano, e comparar com a área total de cada um. Tal informação poderia ser inserida em um *popup*. O próprio Leaflet cria a escala de cores de acordo com os valores da base de dados.

Ressalta-se que, para o trabalho com o pacote Leaflet, é exigido uma base de dados que contenha as coordenadas geográficas das localidades que se deseja plotar. Logo, na configuração desses mapas foi utilizado uma base de dados do tipo “SpatialPolygonsDataFrame” com a função de mesmo nome no R, que contém diversas informações, entre elas, as coordenadas geográficas para a construção do contorno desejado do mapa da Amazônia Legal.

#### 4 | APPS PARA ANALISAR O DESMATAMENTO E FAZER PREVISÕES

A próxima etapa é criar ferramentas que possibilitem ao usuário extrair informações da base para realizar a tomada de decisões. Para tanto, foi criado um *app* com o pacote Shiny, que utiliza a base de dados para gerar um gráfico de barras



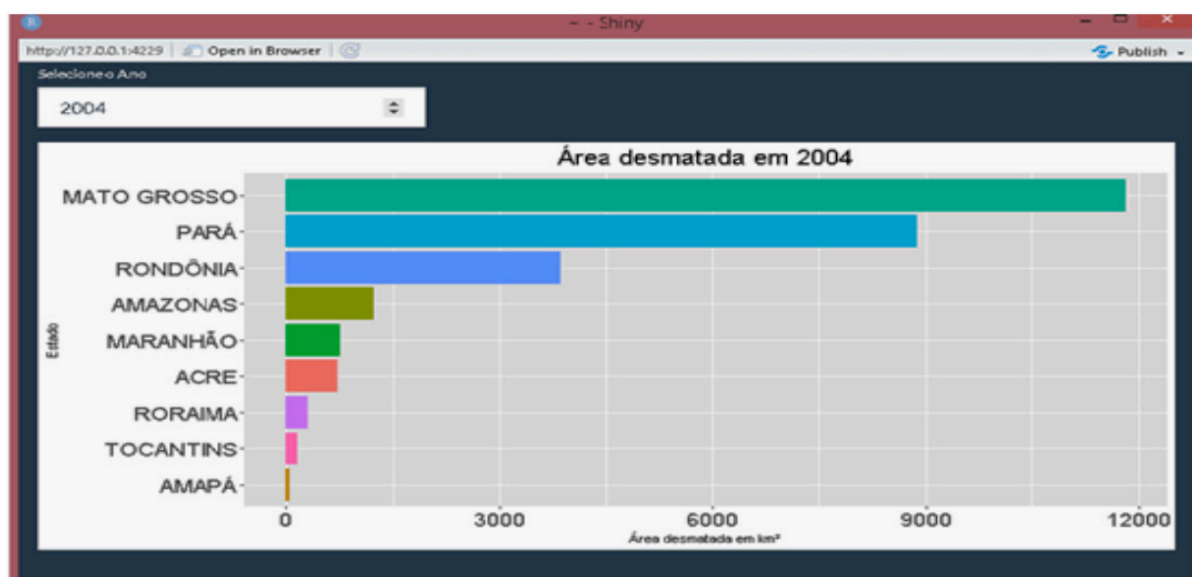
para cada ano. O usuário pode escolher o ano que quer analisar e, automaticamente, é gerado esse gráfico.

O *app* construído possui uma entrada, onde o usuário escolhe o ano que deseja analisar e uma saída que apresenta um gráfico de barras com as áreas de desmatamento na Amazônia Legal daquele ano. A principal vantagem é a possibilidade de colocá-lo em um navegador para assim compartilhar as informações e análises desejadas.

É possível acrescentar outras informações a esse *app* incluindo, por exemplo, medidas de estatística descritiva e de dispersão que podem auxiliar na implantação de políticas públicas para combater o desmatamento. A Figura 7 mostra a interface desse *App* com os gráficos dos anos de 2004, 2005 e 2019.

O *App* da Figura 8 tem a proposta de construir um modelo de previsão com base nos dados da Amazônia Legal. Para isto, foi usada a função *auto.arima()* presente no pacote *Forecast*.

O usuário seleciona o estado que deseja analisar e, automaticamente, são geradas duas saídas. A primeira é um gráfico representando a série temporal do Estado selecionado e a segunda também é um gráfico, mas para o Modelo de Previsão. Nesse modelo, os três últimos pontos em azul representam as áreas de desmatamento previsto para 2019, 2020 e 2021. Optou-se por considerar o ano de 2019 como medida comparativa para os resultados gerados pelo modelo, comparando-o com o resultado real. As áreas destacadas em cinza representam o intervalo de confiança padrão para o modelo utilizado.



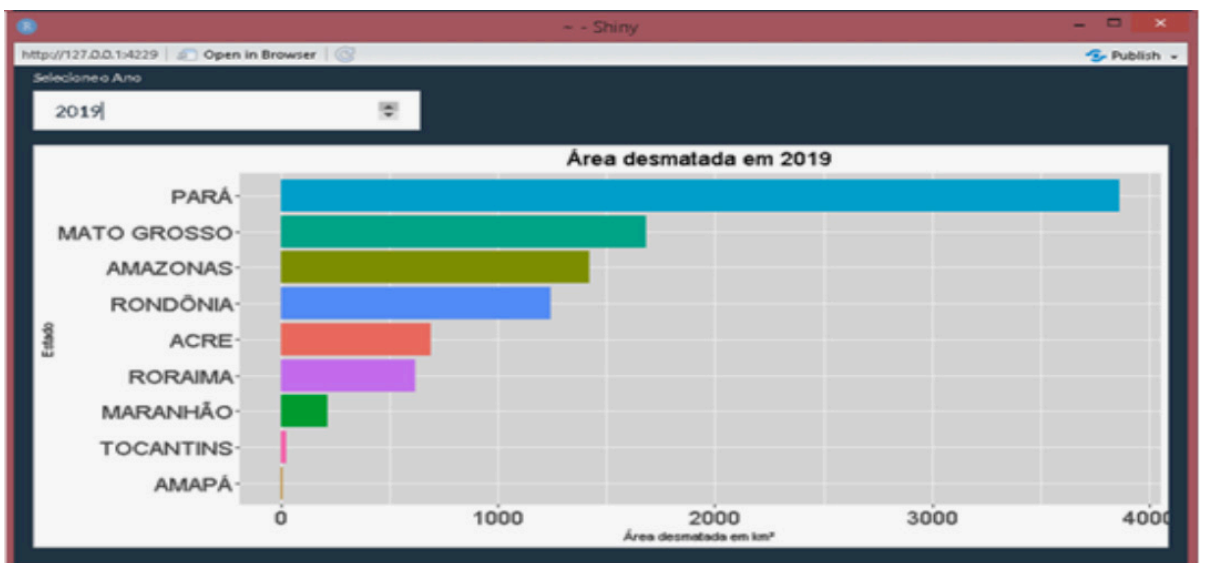
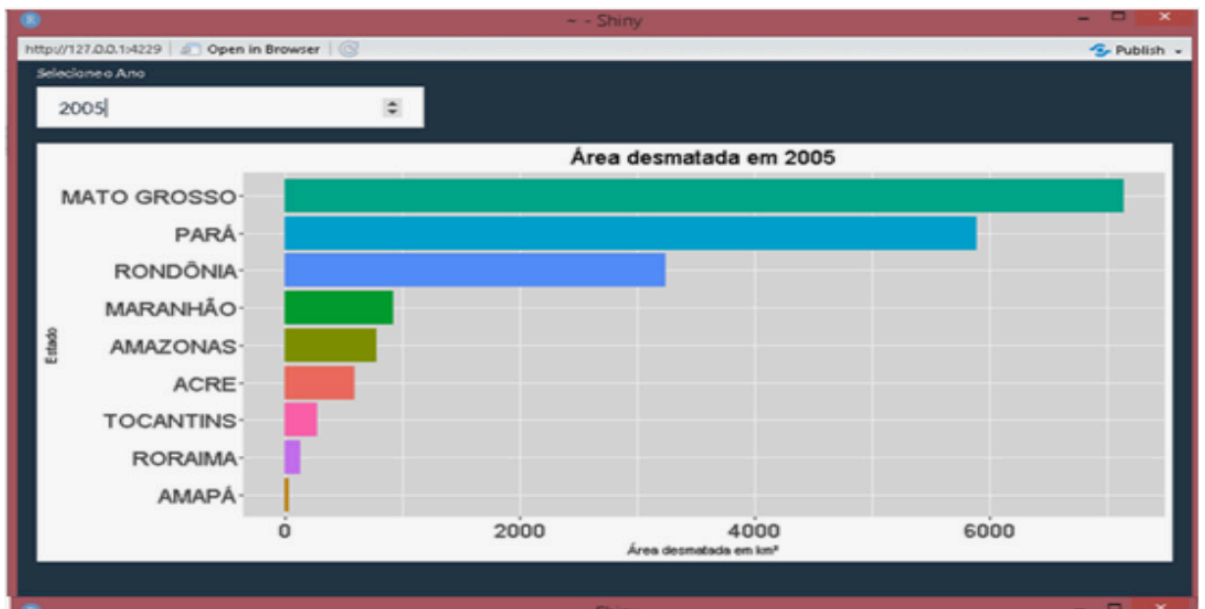


Figura 7: App com a tendência anual

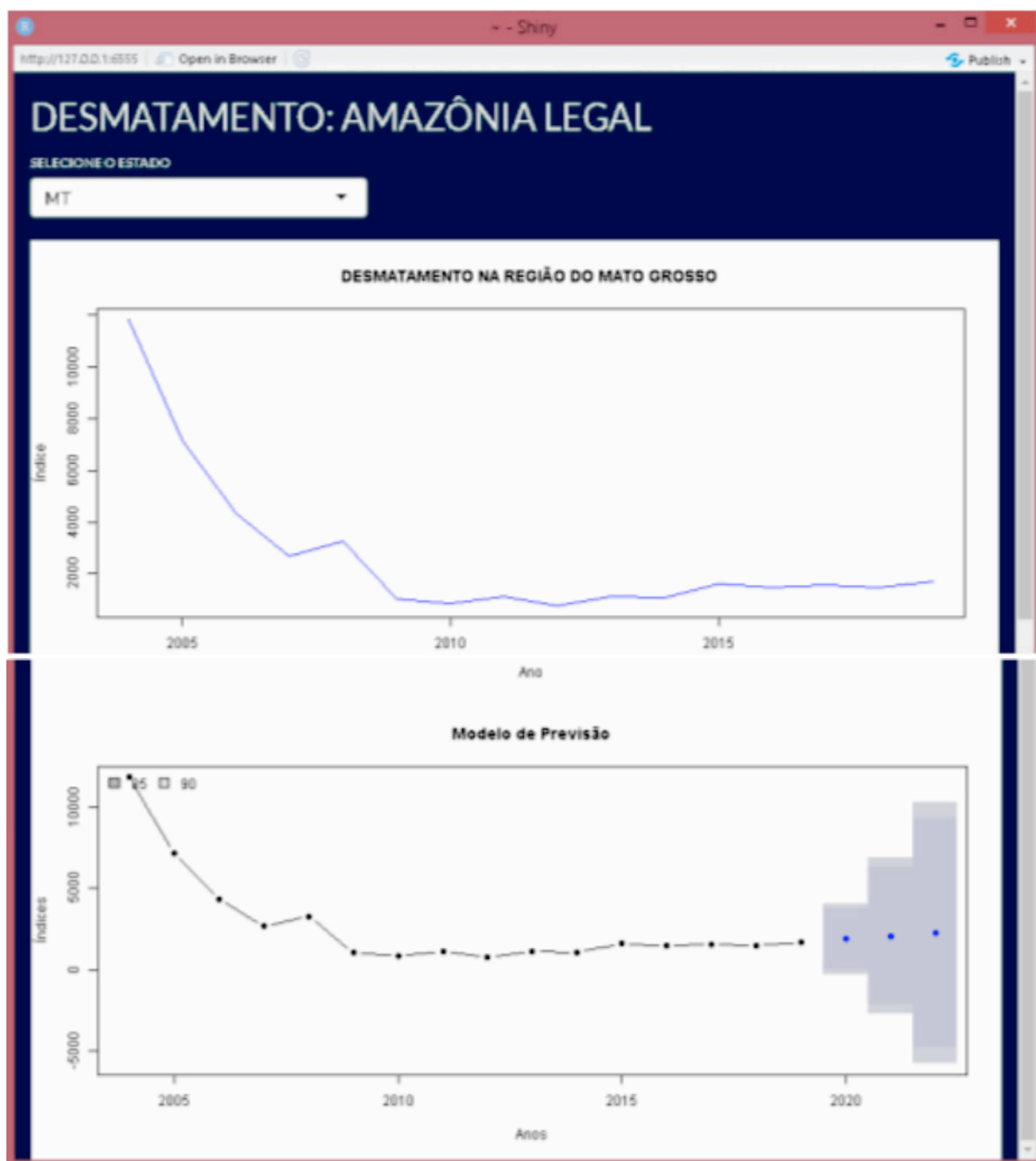


Figura 8: App com o modelo de previsão

A área de desmatamento prevista para o ano de 2019 foi de 1419 km<sup>2</sup>, conforme podemos observar na Tabela 3, (o valor real divulgado pelo PRODES foi de 1685 km<sup>2</sup>), bem como, para 2020 é esperado uma área desmatada de 1348 km<sup>2</sup> e 1277 km<sup>2</sup> para 2021, de acordo com o modelo. Segundo o governo do estado de Mato Grosso, aproximadamente 50% de seu território é composto por florestas, ou seja, 451.678,954 km<sup>2</sup>. Considerando o período analisado (2004 a 2018) e os três anos de projeção do modelo, (2019 a 2021), um total de 45.692 km<sup>2</sup> de área terá sido desmatada, o que equivale a mais de 10% de toda a área florestal deste estado e, tudo isso, em apenas 18 anos.

Ano	Área_desmatada
2019	1419
2020	1348
2021	1277

Tabela 3: Previsões do modelo

Com esse *App* é possível analisar as séries temporais de todos os estados e verificar a tendência de desmatamento para os próximos anos.

## 5 | CONSIDERAÇÕES FINAIS

O desmatamento na Amazônia tem sido tema de reportagens e discussões no mundo inteiro ao longo dos últimos anos. Os avanços computacionais permitem a utilização de ferramentas de visualização interativa em conjunto com os cálculos estatísticos descritivos ou modelos estatísticos de previsão. Essa integração mostra-se útil para analisar e entender diferentes tipos de problemas.

O software R ganha destaque entre as ferramentas computacionais para realizar apresentações gráficas, análises e previsões. É possível organizar, analisar e tomar decisões com base nos dados. A variedade de pacotes disponíveis permite incrementar o processo de análise por meio de modelos mais sofisticados e que, porventura, requerem um esforço mental adicional. Dessa forma, problemas essenciais como a questão do desmatamento podem ser discutidos e analisados de forma mais consistente.

Realizou-se uma análise descritiva inicial, entendendo o comportamento dos dados, utilizando diferentes tipos de representações gráficas e realizando diferentes análises, mostrando o comportamento anual, o comportamento por região e o agregado geral do nosso período. Isso também permitiu analisar a variabilidade dos dados, além de fazer previsões com base em modelos matemáticos para os próximos anos.

Mais que uma ferramenta para a construção de dashboards, os *Apps* construídos permitem a interação entre o desenvolvedor e os usuários, de tal forma que, o usuário consiga, de maneira simples, ler e analisar os dados apresentados. A proposta leva em consideração uma situação específica, porém tal ideia pode ser estendida para diferentes tipos de problemas em diferentes segmentos. Esses *Apps* podem auxiliar na conscientização e fomentar a discussão desses casos, levando a diferentes análises, por pessoas com diferentes conhecimentos, fomentando assim a busca por soluções.

## 6 | MATERIAL DE APOIO

Os scripts para executar os *apps* para análise de desmatamento e demonstração gráfica de suas séries temporais podem ser baixados no link:

<http://doi.org/10.13140/RG.2.2.28851.78881>

### REFERÊNCIAS

BUSSAB, W.O. MORETTIN, P.A. **Estatística básica**. São Paulo: Atual Editora, 1988.

CHANG, W. CHENG, J. ALLAIRE, J. XIE, Y. McPHERSON, J. (2019). shiny: Web Application Framework for R. **R package version 1.3.2**. <https://CRAN.R-project.org/package=shiny>.

CHENG, J. BHASKAR, K. XIE, Y. Leaflet: create interactive web maps with the JavaScript 'Leaflet' Library. **R package version 2.0.2**. <https://CRAN.R-project.org/package=leaflet>.

FEARNSIDE, P. M. **Desmatamento na Amazônia: dinâmica, impactos e controle**. Acta Amaz. [online]. 2006, vol.36, n.3, pp.395-400.

GOVERNO DE MATO GROSSO. **Geografia**. Disponível em: <<http://www.mt.gov.br/geografia>>. Acesso em 27 de nov. de 2019.

GROLEMUND, G. WICKHAM, H. **R para data Science**. Rio de Janeiro: Alta Books, 2019.

HYNDMAN, R. ATHANASOPOULOS, G. BERGMEIR, C. CACERES, G. CHHAY, L. O'HARA-WILD, M. PETROPOULOS, F. RAZBASH, S. WANG, E. YASMEEN, F. (2019). `_forecast`: Forecasting functions for time series and linear models\_. **R package version 8.7**, <<http://pkg.robjhyndman.com/forecast>>

Monitoramento do desmatamento da floresta amazônica brasileira por satélite. **Inpe**, 2019. Disponível em: <[www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes](http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes)>. Acesso em 11 de nov. de 2019.

O'GRADY,S.The RedMonk programming language ranking. **Red Monk**. Disponível em: <<https://redmonk.com/sogrady/2018/08/10/language-rankings-6-18/>>. Acesso em: 19 de nov. de 2019.

R Development Core Team (2019) R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<https://cran.r-project.org/>>

RStudio, 2019 < <https://www.rstudio.com/products/rstudio/>>

SAS. Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/analises-preditivas.html](https://www.sas.com/pt_br/insights/analytics/analises-preditivas.html). Acesso em: 26 de nov. de 2019

SCHMULLER, J. **Análise estatística com R para leigos**. Rio de Janeiro: Alta Books, 2019.

SCHMULLER, J. **Projetos em R para leigos**. Rio de Janeiro: Alta Books, 2019.

WICKHAM, H. ggplot2: Elegant Graphics for Data Analysis. **Springer-Verlag**. New York, 2016.

WICKHAM, H. tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>

## AVALIAÇÃO DE GRADIENTE PEDOAMBIENTAL USANDO ANÁLISE DE COMPONENTES PRINCIPAIS (PCA) NA ANTÁRTICA MARÍTIMA

Data de aceite: 12/02/2020

### Daniela Schmitz

Departamento de Solos, Universidade Federal de Viçosa  
Viçosa, MG

### Pedro Manuel Villa

Departamento de Engenharia Florestal, Universidade Federal de Viçosa  
Viçosa, MG

### Carlos Ernesto G.R. Schaefer

Departamento de Solos, Universidade Federal de Viçosa  
Viçosa, MG

### Márcio Rocha Francelino

Departamento de Solos, Universidade Federal de Viçosa  
Viçosa, MG

**RESUMO:** A Antártica é o principal ecossistema terrestre dominado por comunidades de plantas não-vasculares, que são restritas às poucas áreas livres de gelo. Estas comunidades estão distribuídas ao longo de gradientes ambientais, como por exemplo em diferentes pedoambientes. A análise de componentes principais (PCA) é um dos métodos de estatística multivariada amplamente usado para identificar e estudar gradientes ecológicos. Assim, o objetivo deste trabalho é demonstrar a aplicabilidade do uso

da PCA para a identificação de gradientes pedoambientais através da variabilidade das características físico-químicas do solo em uma área livre de gelo da Antártica Marítima. Foram selecionadas quatro áreas com distintas coberturas de comunidades de musgos e avaliadas as propriedades do solo. Foram usados os seguintes pacotes do R: “FactoMineR” que é utilizado para gerar a PCA; “factoextra” é usado para construir os gráficos da PCA; e “ggplot2” para editar os gráficos. Além disso foi usado o pacote “corrplot” para a análise de correlação das variáveis do solo com eixos da PCA. Uma alta variabilidade nas propriedades do solo foi observada ao longo do gradiente pedoambiental. A PCA determinou a separação dos quatro grupos, formando um marcado gradiente pedoambiental, baseado tanto nos atributos químicos quanto físicos do solo. A PCA é uma importante ferramenta estatística para separar grupos (ambientes) e agrupar as variáveis (propriedades do solo) que estejam atuando fortemente sobre eles. Sendo dessa forma, o passo inicial muito útil em estudos que objetivam avaliar o efeito de um gradiente ambiental sobre a vegetação.

**PALAVRAS-CHAVE:** análise de gradiente; análise de correlação; análise multivariada; PCA r script; propriedades do solo.



## EVALUATION OF PEDOENVIRONMENTAL GRADIENT USING PRINCIPAL COMPONENT ANALYSIS (PCA) IN MARITIME ANTARCTICA

**ABSTRACT:** Antarctica is the main terrestrial ecosystem dominated by non-vascular plant communities, which are restricted to a few ice-free areas. These communities are distributed along environmental gradients, such as different pedoenvironments. Principal component analysis (PCA) is one of the multivariate statistical methods widely used to identify and study ecological gradients. Thus, the objective is to demonstrate the applicability of the use of PCA for the identification of pedoenvironmental gradients through the variability of soil physico-chemical characteristics in Maritime Antarctica ice-free areas. Four areas with different plant cover were selected and soil properties evaluated. The following R packages were used: “FactoMineR” which is used to generate the PCA; “Factoextra” is used to build PCA graphics, and “ggplot2” to edit the graphics. Also, the corrplot package was used for the correlation analysis of soil variables with PCA axes. High variability in soil properties was observed along the pedoenvironmental gradient. The PCA determined the separation of the four groups, forming a marked pedoenvironmental gradient, based on both chemical and physical soil attributes. PCA is an important statistical tool for separating groups (environments) and grouping variables (soil properties) that are strongly acting on them. Thus, the initial step is very useful in studies that aim to evaluate the effect of an environmental gradient on vegetation.

**KEYWORDS:** gradient analysis; correlation analysis; multivariate analysis; PCA R script; soil properties.

### 1 | INTRODUÇÃO

Os solos da antárticos são resultado de processos pedológicos que ocorrem de forma lenta nas poucas áreas livres de gelo (SCHAEFER et al., 2004). Na Antártica, as variações climáticas e o material parental são fatores-chave para entender a distribuição de solos e suas propriedades (Daher, 2019). Um contrastante desenvolvimento do solo pode ser observado em fina escala em áreas periglaciais, apresentando diferentes características físico-químicas que refletem a influência de variáveis ambientais como microclima, geologia, geomorfologia, drenagem, deposição de guano e cobertura vegetal (THOMAZINI et al., 2018). Desta forma, essa variabilidade nas características do solo também é responsável na distribuição da riqueza e composição de espécies não vasculares, e na cobertura vegetal ao longo de um gradiente pedoambiental na Antártica Marítima (SCHMITZ et al., 2020). No entanto, para realizar estudos com o objetivo de entender como a vegetação responde ao desenvolvimento do solo, é necessário primeiramente, reconhecer e definir a existência de um gradiente pedoambiental.

Em estudos ambientais com muitas amostras e variáveis, sejam elas físico-

químicas ou biológicas, uma das formas de avaliar integradamente um complexo conjunto de dados é a aplicação da estatística multivariada, onde as variáveis podem ser analisadas simultaneamente (BERNARDI et al., 2009). A análise de componentes principais (PCA) é uma análise multivariada amplamente usada na ecologia e que possibilita investigações com um grande número de dados disponíveis (VINCINI, 2005). Permite interpretar a estrutura de um conjunto de dados multivariados a partir das respectivas matrizes de variâncias-covariâncias ou de correlações entre as variáveis de estudo (BERNARDI et al. 2009). Possibilita também, a identificação das medidas responsáveis pelas maiores variações entre os resultados, sem perdas significativas de informações.

Esta técnica pode ser utilizada para geração de índices e agrupamento de indivíduos ou variáveis ambientais. A análise agrupa ambientes (ou indivíduos, ou o que for o objeto de estudo) de acordo com suas variações, representado pela variação do conjunto de características que define o ambiente (ou objeto de estudo), ou seja, a técnica agrupa ou separa os ambientes de uma área segundo a variação das suas características (HONGYU, et al., 2015). Para evitar conclusões indesejáveis, o pesquisador deve ter, porém, um bom conhecimento a respeito das variáveis em questão, pois a análise multivariada é apenas uma forma exploratória de classificação e ordenação (BERNARDI et al., 2009).

Neste caso, usaremos dados de um levantamento de solo de 60 parcelas, representando quatro ambientes situados em Rip Point, na Ilha Nelson, onde os parâmetros utilizados são as variáveis químicas e físicas do solo. O objetivo deste trabalho é demonstrar a aplicabilidade do uso da Análise de Componentes Principais para a identificação de gradientes pedoambientais através da variabilidade das características físico-químicas do solo em uma área livre de gelo da Antártica Marítima (AM). E assim, descrever de forma prática e detalhada o uso desta ferramenta estatística para a reprodução e uso na análise multivariada de dados ambientais e ecológicos.

## **2 | METODOLOGIA**

### **2.1 Área de estudo e coleta dos dados de solo**

Este estudo foi conduzido na área de Rip Point, na Ilha Nelson, Antártica Marítima. É uma área periglacial de aproximadamente 3 km de comprimento por 1.5 km de largura (PUTZKE et al. 1998). Foram selecionadas quatro áreas com distintas coberturas de comunidades de musgos. Em cada área foram levantadas 15 parcelas de 20x20cm, totalizando 60 parcelas amostradas.

Para medir as propriedades do solo, foi coletada uma amostra simples de solo superficial (0-10cm) em cada parcela. As amostras foram enviadas para análise

de rotina do laboratório de solos da UFV seguindo os protocolos padronizados (EMBRAPA, 1997). As seguintes propriedades do solo foram determinadas: pH (H<sub>2</sub>O), P permutável, K<sup>+</sup>, Ca<sup>2+</sup>, Na, Mg<sup>2+</sup>, Fe, Cu, Mn, Zn, Al<sup>3+</sup>, acidez trocável (H + Al), matéria orgânica (MO), soma de bases trocáveis (SB); capacidade potencial de troca catiônica (T), porcentagem de saturação por bases (V), e a textura do solo, como teores de areia grossa (A\_grossa), areia fina (A\_fina), silte e argila.

## 2.2 Organização da planilha para análise de dados no programa R

Após a obtenção das variáveis físico-químicas do solo, é necessário organizá-las de forma sistemática em uma planilha Excel. A seguir é mostrado como as variáveis devem ser sistematizadas na planilha.

Pedoambiente	Parcela	pH	P	K	Na	Ca	MO	...	Areia	...
P1	1	6.51	101.7	248	646.2	7.4	0.65	...	0.771	...
P1	2	6.5	170.7	336	935.4	6.21	0.26	...	0.712	...
P1	3	6.12	85.9	258	496.7	7.13	1.7	...	0.821	...
P1	4	6.41	124.1	316	935.4	5.04	0.52	...	0.665	...
...	...	...	...	...	...	...	...	...	...	...
P2	21	5.91	56.6	146	397.0	28.43	1.57	...	0.428	...
P2	22	5.99	221.1	144	397.0	37.58	1.83	...	0.501	...
P2	23	6.44	86.3	188	486.7	27.34	0.78	...	0.523	...
P2	24	6.2	121.7	126	755.9	22.38	1.3	...	0.401	...
...	...	...	...	...	...	...	...	...	...	...
P3	41	6.4	132.5	62.3	486.7	19.95	0.26	...	0.751	...
P3	42	6.3	71.9	47.1	456.8	14.89	0.52	...	0.656	...
...	...	...	...	...	...	...	...	...	...	...
P4	61	6.5	485.8	152	297.2	16.02	3.13	...	0.610	...
P4	62	6.3	246.6	146	267.3	12.06	4.7	...	0.694	...
...	...	...	...	...	...	...	...	...	....	...

Tabela 1. Organização dos pedoambientes, parcelas e variáveis químicas e físicas do solo na planilha Excel para posterior análise no software R.

As duas primeiras colunas correspondem aos pedoambientes (P1, P2, P3 e P4) e as parcelas (1, 2, 3, 4, ..., 60), respectivamente. As colunas seguintes são as variáveis que pretendemos analisar, neste caso, atributos químicos (pH, P, K, etc.) e físicos do solo (areia, silte e argila). Após a sistematização dos dados na planilha é necessário salvá-la em formato txt, observando sempre que as vírgulas (,) devem ser substituídos por ponto (.).

## 2.3 Pacote do R e comandos de análise de dados (script)

Para a análise de dados, construção e edição dos gráficos PCA, foram utilizados seis pacotes no software R: ‘FactoMineR’ que é utilizado para resumir, visualizar e descrever conjuntos de dados quando as variáveis são categóricas e análise de fatores múltiplos quando variáveis são estruturadas em grupos (HUSSON et al., 2017); ‘factoextra’ fornece algumas funções para extrair e visualizar a saída de análises multivariadas de dados (incluindo PCA), também contém funções para simplificar algumas etapas de análise de clustering e fornece uma visualização de dados elegante do ‘ggplot2’ (KASSAMBARA, 2017); ‘ggplot2’ é um pacote do R usado para criar gráficos estatísticos de forma declarativa, baseado em mapeamento de atributos estéticos de forma geométricas (WICKHAM & CHANG, 2019); ‘corrplot’ que faz uma exibição gráfica de uma matriz de correlação das variáveis analisadas no PCA, incluindo a escolha de cores, rótulos de texto, etiquetas coloridas, layout (WEI, 2017).; ‘ggrepel’ que ajuda a evitar a sobreposição de títulos e rótulos nos gráficos (SLOWIKOWSKI et al., 2019); e ‘vegan’ que é utilizado para métodos de ordenação, análise de diversidade e outras funções para vegetação e ecologia de comunidades (OKSANEN et al., 2019).

## 2.4 Detalhes do script

A seguir serão descritos todos os comandos utilizados, desde o carregamento dos dados no R até o produto final das análises, gráfico da PCA e valores de análises de correlação de variáveis.

Primeiramente é preciso selecionar endereço do arquivo ou a pasta de trabalho onde se encontram os dados/planilhas a serem analisados. Nessa pasta também serão salvos todos os gráficos gerados nas análises.

```
setwd (“~/endereço do arquivo/”)
```

O comando a seguir é utilizado para discriminar a planilha em formato txt que será o objeto de análise.

```
dados <- read.table(“matriz.txt”, h=T, sep=“\t”)
```

Para identificar, consultar e explorar os dados da planilha usamos os seguintes comandos:

```
table(dados$Group)
```

```
dados
```

```
dim(dados)
```

```
names(dados)
```

```
head(dados)
```

```
summary(dados[, -1])
```

A função *dim* é usada para visualizar as dimensões da planilha, neste caso, 60 parcelas e 21 variáveis. Enquanto que a função *names* é usada para mostrar os

nomes das variáveis (primeira linha) que estamos utilizando nas análises. A função *summary* é utilizada para mostrar os valores mínimo, médio, mediano, máximo, 1º e 3º quartis de cada variável.

Antes de começar a análise dos dados, é necessário a instalação dos pacotes, que podem ser instalados com a seguinte função:

```
install.packages ("nome do pacote")
```

Após instalar o pacote, é necessário “carregar” o pacote sempre que você abrir o R, utilizando a função *library*.

```
library (FactoMineR)
```

```
library (factoextra)
```

```
library (ggplot2)
```

Onde o pacote ‘FactoMineR’ é utilizado para rodar a PCA; o pacote ‘factoextra’ é usado para construir os gráficos da PCA; e ‘ggplot2’ para editar os gráficos.

Podemos identificar a matriz de dados diretamente no função PCA:

```
Gradiente <- PCA(dados[,3:21], graph = FALSE)
```

Neste caso, demos a ela o nome de “Gradiente”, e então selecionamos todas as linhas (,) e as colunas (3 a 21) que representam as variáveis químicas e físicas do solo usadas na análise. Ou também podemos criar objetos para definir a matriz de dados, como por exemplo, chamaremos de “matriz” todos os parâmetros selecionados.

```
matriz <- dados[, 3:21]
```

```
matriz
```

E então na função para rodar a PCA devemos colocar a matriz de interesse:

```
Gradiente <- PCA(matriz, graph = FALSE)
```

Em seguida exploramos diferentes opções de gráficos para a PCA, para observar as contribuições dos eixos, relação de vetores, distribuição de grupos, identificar gradientes.

O “graph of individuals” indica formação de grupos com identificação de parcelas: (Fig 1a).

```
fviz_pca_ind(Gradiente)
```

O “graph of variables”, identifica os vetores que correspondem a cada parâmetro (Fig 1b):

```
fviz_pca_var(Gradiente)
```

Para indicar o nível de correlação dos vetores com escala de cores:

```
fviz_pca_var(Gradiente, col.var = "cos2", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), repel = TRUE)
```

Onde a função *fviz\_pca\_var* é a função que identifica os vetores que correspondem a cada parâmetro; o argumento *col.var* dá cores as variáveis e “cos2” indica que as cores para as variáveis são selecionadas automaticamente controladas por qualidade de representação; e *gradient.cols* significa que será seguido um gradiente entre as cores selecionadas.

Para um gráfico PCA Biplot, isto é, usando parcelas e variáveis (Fig. 1c):  
`fviz_pca_biplot(Gradiente, geom = "text")`

Onde o argumento *geom* é um texto especificando a geometria a ser usada para o gráfico. Os valores permitidos são a combinação de *c*("point", "arrow", "text"). Use "point" (para mostrar apenas pontos); "text" para mostrar apenas etiquetas; *c* ("point", "text") ou *c* ("arrow", "text") para mostrar os dois tipos. Ou também da seguinte forma:

```
fviz_pca(Gradiente)
```

Para definir e diferenciar grupos adicionamos o argumento *habillage* que é uma variável opcional para colorir as observações por grupos. O valor padrão é "none":

```
fviz_pca_ind(Gradiente, label="none", habillage=dados$Group)
```

Se *X* é um objeto PCA, o *habillage* também pode especificar a variável qualitativa suplementar (por seu índice ou nome) a ser usada para colorir indivíduos por grupos:

```
fviz_pca(Gradiente, habillage=dados$Group)
```

Para visualizar a contribuição de cada eixo da PCA:

```
fviz_screplot(Gradiente, ncp=10)
```

E incluir as porcentagens de contribuição de cada eixo:

```
fviz_eig(Gradiente, addlabels = TRUE)
```

Para obter os valores de correlação por eixo para cada variável:

```
Gradiente$var$coord
```

Para obter uma representação gráfica dessa correlação por eixo para cada variável, instalamos o pacote "corrplot" e prosseguimos com o seguinte comando:

```
library(corrplot)
```

```
corrplot(Gradiente$var$coord, is.corr=F)
```

Para construir um gráfico completo da PCA de solos que inclui todos os parâmetros usamos o pacote "ggplot2", que já foi instalado anteriormente, que tem funções que possibilitam fazer ajustes de edição de gráfico, e adicionar manualmente cores, formas e valores de eixos:

```
fviz_pca_biplot(Gradiente, habillage=dados$Group, label = "var", col.var = "red",  
alpha.var = "cos2", repel = TRUE, addEllipses = TRUE) + theme_minimal() + theme_  
gray(base_size = 14) + labs(title="", x = "PCA1 (33.2%)", y = "PCA2 (28.1%)") + scale_  
color_manual(values = c("red3", "gold2", "royalblue", "blue"), name = "Pedoambiente",  
limits=c("P1", "P2", "P3", "P4")) + scale_shape_manual(values=c(15, 16, 17, 14),  
name = "Pedoambiente", limits=c("P1", "P2", "P3", "P4")) + guides(fill = FALSE)
```

Onde o argumento *label* é um texto especificando os elementos a serem rotulados. O valor padrão é "all". Os valores permitidos são "none" ou a combinação de *c*("ind", "ind.sup", "quali", "var", "quanti.sup"). "Ind" pode ser usado para rotular apenas indivíduos ativos. "Ind.sup" é para indivíduos complementares. "Quali" é para variáveis qualitativas complementares. "Var" é para variáveis ativas. "Quanti.



sup” é para variáveis suplementares quantitativas.

O argumento *col.var* é usado para definir a cor dos vetores das variáveis, que neste caso escolhemos “red” (vermelho).

O argumento *alpha.var* controla a transparência das cores das variáveis, respectivamente. O valor pode variar de 0 (transparência total) a 1 (sem transparência). O valor padrão é 1. Os valores possíveis incluem também: “cos2”, “contrib”, “coord”, “x” ou “y”. Nesse caso, a transparência das cores individuais / variáveis é controlada automaticamente por suas qualidades (“cos2”), contribuições (“contrib”), coordenadas ( $x^2 + y^2$ , “coord”), valores x (“x”) ou valores y (“y”). Para usar isso, verifique se *habillage* = “none”.

O argumento *repel* é para evitar sobreposição de rótulos que indentificam variáveis.

O argumento *addEllipses* se “TRUE” (verdadeiro), desenha elipses ao redor dos indivíduos.

O argumento *theme* é uma cadeia de caracteres ou expressão indicando o título do gráfico. Se NULL, o título não é mostrado. O nome do objeto de escala ou o nome especificado em *labs()* é usado para o título. Neste caso, usado para nomear os eixos 1 e 2 da PCA com os valores já obtidos acima.

O argumento *scale\_color\_manual* é usado para selecionar manualmente as cores para cada pedoambiente; *name* é usado para dar nome a legenda, que em nosso caso é chamado “Pedoambiente” e *limits* para definir quais são cada um deles e como são chamados (P1, P2, P3 e P4). O argumento *scale\_shape\_manual* é usado para selecionar manualmente as formas de para cada pedoambiente (cada número representa uma forma); repetindo *names* e *limits* como anteriormente para concatenar cores e formas para cada pedoambiente e finalizando com o argumento (*fill* = FALSE), para anular função de preenchimento e evitar repetir legendas no gráfico.

Para salvar a imagem do gráfico na pasta de trabalho dos arquivos é usada a função *ggsave* do ggplot2, inserindo o nome e formato, a dimensão (“width” e “height”) e a resolução desejada (“dpi”):

```
ggsave(“nome da PCA.jpeg”, width = 8, height = 5, dpi = 300)
```

Baseado no exemplo anterior, para adicionar a numeração das parcelas, instalamos o pacote “ggrepel” e os comandos a seguir:

```
library(ggrepel)
```

```
fviz_pca_biplot(Gradiente, habillage=dados$Group, label = “var”, col.var = “red”,  
alpha.var = “cos2”, repel = TRUE, addEllipses = TRUE) + theme_minimal() + theme_  
gray(base_size = 14) + labs(title=”, x = “PCA1 (33.2%)”, y = “PCA2 (28.1%)”) + scale_  
color_manual(values = c(“red3”, “gold2”, “royalblue”, “blue”), name = “Pedoambiente”,  
limits=c(“P1”, “P2”, “P3”, “P4”)) + scale_shape_manual(values=c(15, 16, 17, 14),
```

```
name = "Pedoambiente", limits=c("P1", "P2", "P3", "P4")) + guides(fill = FALSE) +
geom_text_repel(aes(label = dados$Plot), size = 3, color = "black")
```

Onde foi adicionado o argumento *geom\_text\_repel* que é usado para adicionar um texto diretamente no gráfico, seguido do tamanho (*size*) e cor (*color*) desejados.

Para salvar a imagem na pasta de trabalho é feito o mesmo procedimento já mencionado anteriormente.

```
ggsave("nome da PCA_com_todas_as_parcelas.jpeg", width = 8, height = 5, dpi = 300)
```

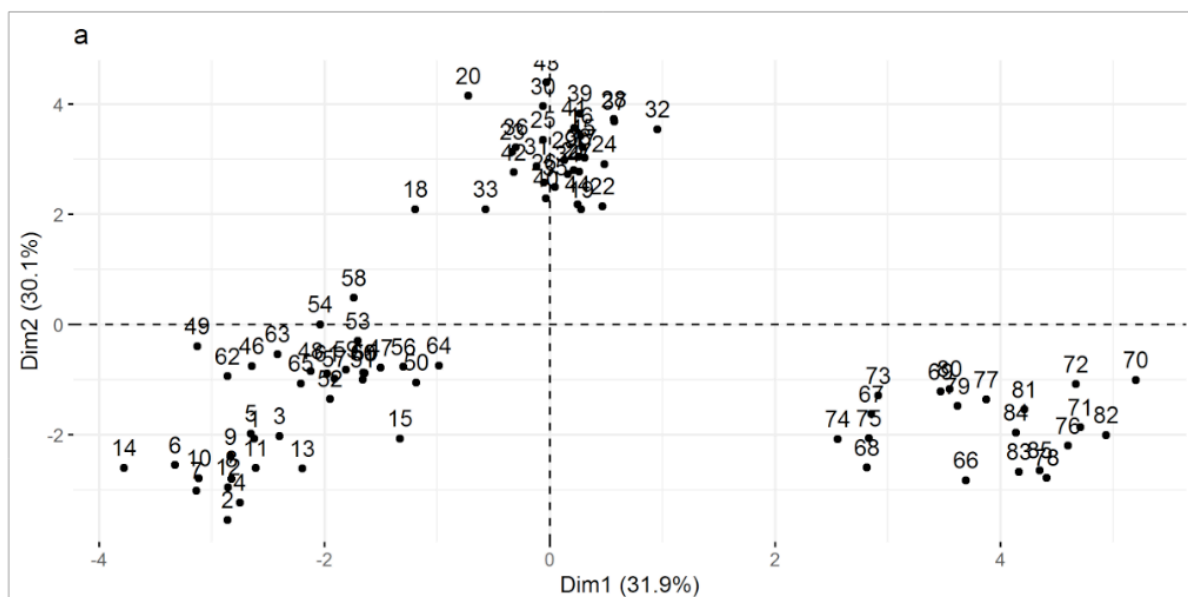
No mesmo gráfico, para usar cores mais discretas (Fig. 2) usamos:

```
fviz_pca_biplot(Gradiente, habillage=dados$Group, label = "var", col.var = "red", alpha.var="cos2", repel=TRUE, addEllipses=TRUE) + theme_minimal() + theme_gray(base_size = 14) +
scale_color_discrete(name = "Pedoambiente", limits=c("P1", "P2", "P3", "P4")) +
scale_shape_manual(values=c(15, 16, 17, 14), name = "Pedoambiente", limits=c("P1", "P2", "P3", "P4")) + guides(fill = FALSE)
```

Onde o argumento *scale\_color\_manual* foi substituído por *scale\_color\_discrete*.

### 3 | RESULTADOS E DISCUSSÃO

A análise de componentes principais (PCA) das propriedades químicas e físicas dos solos permitiu separar e corroborar a separação das quatro áreas de estudo em distintos pedoambientes, formando assim, um acentuado gradiente pedoambiental. Os dois primeiros eixos da PCA explicaram 61.3% da variação dos dados do solo (Fig.2). O primeiro eixo explicou 33.2% da variação dos dados e esteve positivamente correlacionado com silte, MO, N, Zn, P, acidez trocável (H + Al) e negativamente com K<sup>+</sup>, Na e areia grossa (Fig. 3). O segundo eixo da PCA explicou 28.1% da variação dos dados do solo, e esteve positivamente correlacionado com Ca<sup>2+</sup>, Mg<sup>2+</sup>, SB, t e negativamente com pH e P (Fig. 3).



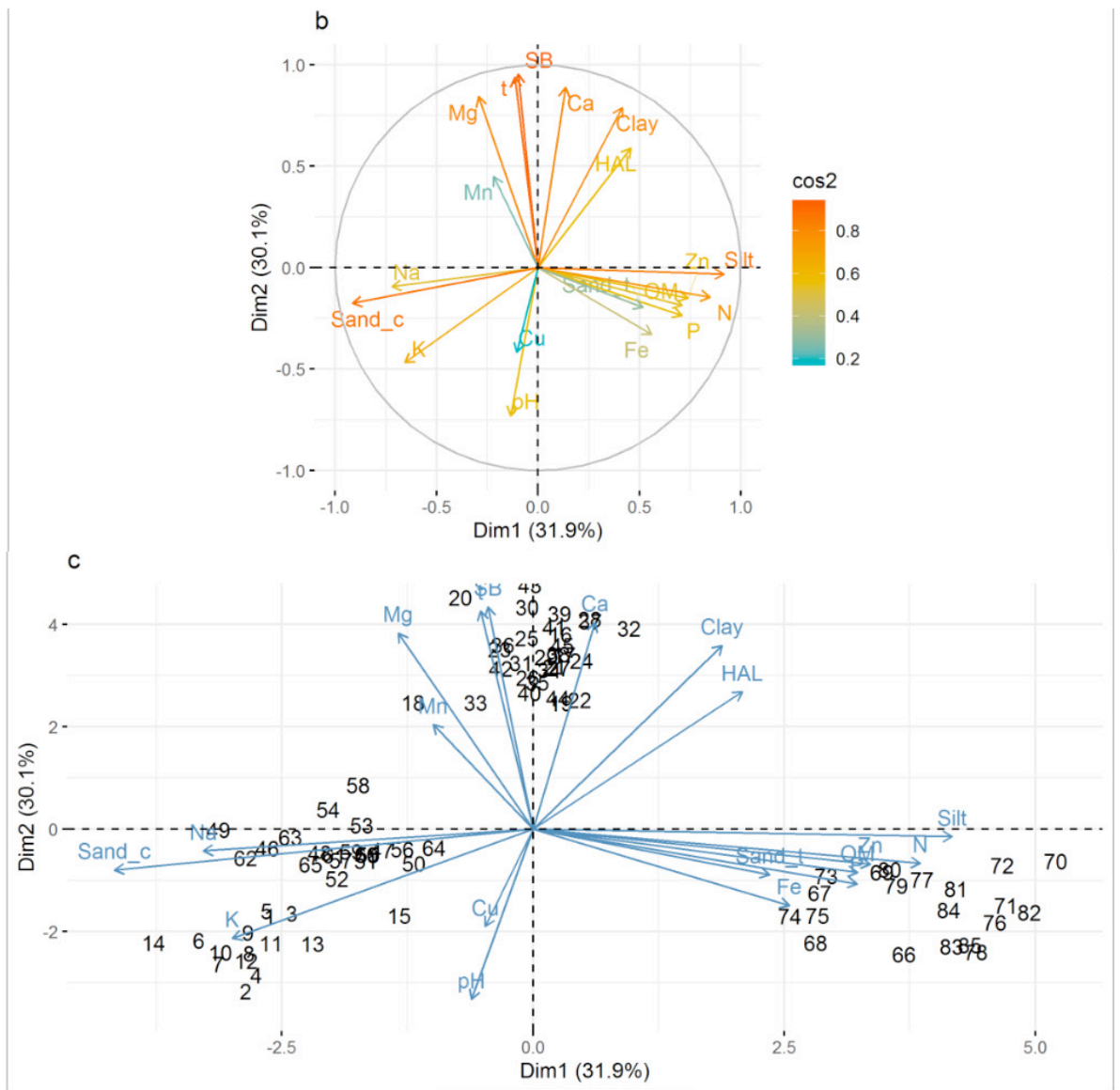


Figura 1. Gráficos da PCA: a) indicando a separação de pedoambientes e identificação das parcelas; b) identificando os vetores que correspondem a cada parâmetro do solo; c) identificando as parcelas e as variáveis.

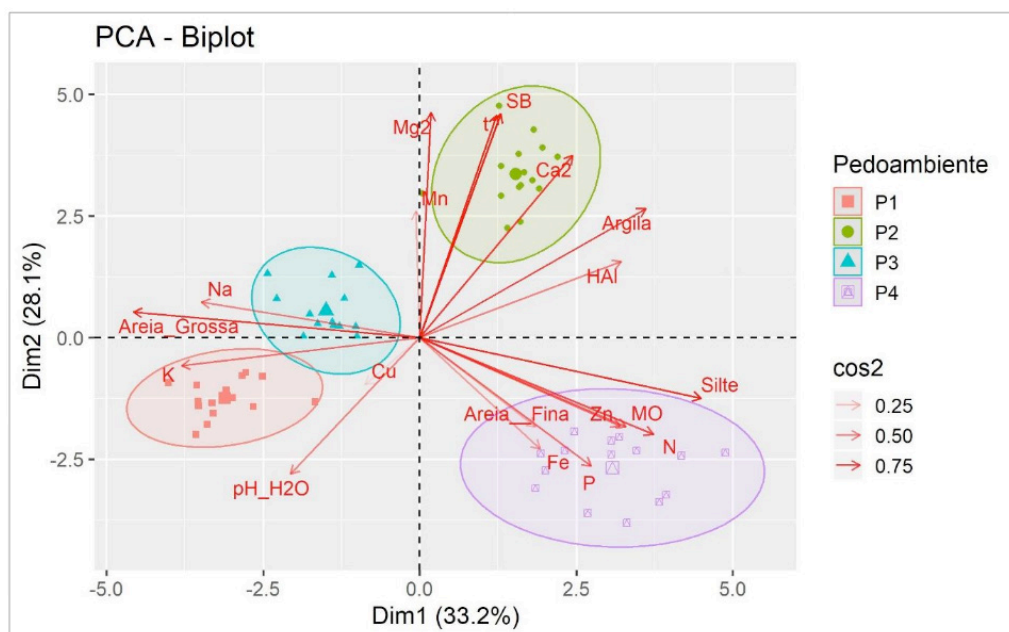


Figura 2. Análise de Componentes Principais (PCA) para os parâmetros do solo nos quatro

Uma alta variabilidade nas propriedades químicas do solo foi observada ao longo do gradiente pedoambiental. O P1 é caracterizado principalmente pelos elevados níveis de K e pH, que são comuns em áreas baixas e recentemente expostas, como terraços marinhos, indicando a ocorrência de lixiviados ricos em K, que são retidos no terraço mais baixo da paisagem (SIMAS et al., 2004). Em P2 são encontrados os maiores valores de  $\text{Ca}^{2+}$ , que podem estar relacionados a uma maior riqueza em fragmentos basálticos no material de origem (SIMAS et al., 2007). Os elevados níveis de  $\text{Na}^+$ , como encontrados em P3, são muito comuns nas áreas mais próximas ao mar, devido a pulverização do spray salino (MICHEL et al., 2006).

As quantidades de areia, silte e argila também se mostraram importantes propriedades do solo para a separação gradiente pedoambiental, condizendo com o encontrado em outras áreas da Antártica Marítima (POELKING et al., 2015). A clara separação entre os pedoambientes permite avaliar os efeitos da filtragem de habitats nas comunidades vegetais em escala local na AM (SCHMITZ et al., 2020).

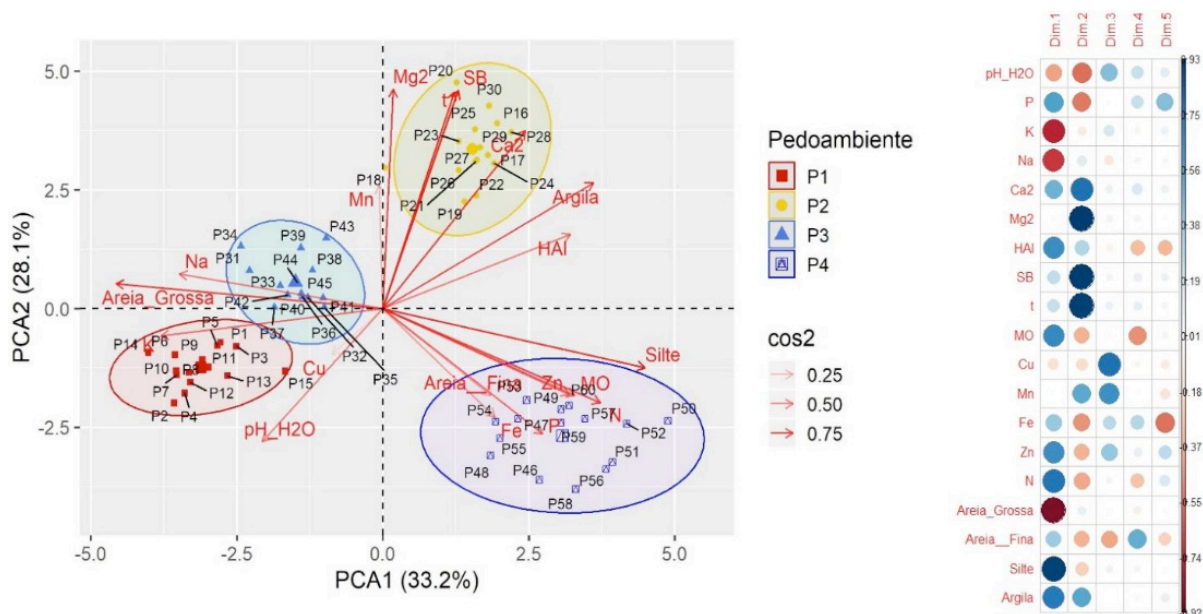


Figura 3. Análise de Componentes Principais (PCA) dos parâmetros do solo com a distribuição das parcelas; e correlação das variáveis do solo com os cinco primeiros eixos da PCA.

#### 4 | CONSIDERAÇÕES FINAIS

A PCA determinou a separação dos quatro grupos, formando um marcado gradiente pedoambiental, baseado tanto nos atributos químicos quanto físicos do solo.

A PCA é uma importante ferramenta estatística para separar grupos (ambientes) e agrupar as variáveis (propriedades do solo) que estejam atuando fortemente sobre eles. Sendo dessa forma, o passo inicial muito útil em estudos que objetivam avaliar o efeito de um gradiente ambiental sobre a vegetação.

## DEPÓSITO DE DADOS

Os arquivos com a matriz dos dados de solo e o script da PCA utilizados podem ser encontrados no link: <http://dx.doi.org/10.13140/RG.2.2.27524.68480>

## REFERÊNCIAS

BERNARDI, J.V.E., LACERDA, L.D., DÓREA, J.G., LANDIM, P.M.B., GOMES, J.P.O., ALMEIDA, R., MANZATTO, A.G., BASTOS, W.R. **Aplicação da análise das componentes principais na ordenação dos parâmetros físico-químicos no Alto Rio Madeira e afluentes, Amazônia Ocidental (2009)**. *Geochimica Brasiliensis* 23(1) 079-090.

DAYER, D., SCHAEFER, C.E.G.R., FERNANDES FILHO, E.I., FRANCELINO, SENRA, E.O. 2019. **Semi-arid soils from a topolithosequence at James Ross Island, Weddell Sea region, Antarctica: Chemistry, mineralogy, genesis and classification**. *Geomorphology* 327: 351-364.

EMBRAPA, 1997. **Manual de Métodos de Análise de Solo**. 2. ed. ver. atualiz. Centro Nacional de Pesquisa de Solos, Rio de Janeiro (212 pp. il., EMBRAPA/CNPS-RJ. Documentos, 1).

HADLEY, W., 2015. **R ggplot2 package**: an implementation of the grammar of graphics. Available at <http://ggplot2.org> , <https://github.com/hadley/ggplot2>.

HONGYU, K., SANDANIELO, V.L.M., OLIVEIRA JR, G.J. **Principal Component Analysis: theory, interpretations and applications (2015)**. *Engineering and Science* 2015, 5:1.

HUSSON, F., JOSSE, J., LE, S., MAZET, J., 2017. **Package FactoMineR**” package Multivariate: Exploratory Data Analysis and Data Mining. <http://CRAN.R-project.org/package=FactoMineR>. RStudio package version 1.0.14.

KASSAMBARA, A. & MUNDT, F. 2017. **Package factoextra**: extract and visualize the Results of Multivariate Data Analyses. R package version 1.0.5.

MICHEL, R.F.M., SCHAEFER, C.E.G.R., DIAS, L., SIMAS, F.N.B., BENITES, V., MENDONÇA, E.S., 2006. **Ornithogenic Gelisols (Cryosols) from Maritime Antarctica: pedogenesis, vegetation and carbon studies**. *Soil Sci. Soc. Am. J.* 70, 1370–1376.

OKSANEN, J., BLANCHET, F.G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P.R., O'HARA, R.B., SIMPSON, G.L., SOLYMOS, P. 2018. **Package Vegan**: Community Ecology Package. R package version 2.0-7.

POELKING, E.L., SCHAEFER, C.E.G.R., FERNANDES FILHO, E.I., DE ANDRADE, A.M., SPIELMANN A.A., 2015. **Soil-landform-plant-community relationships of a periglacial landscape on Potter Peninsula, maritime Antarctica**. *Solid Earth*, 6, 583-594. doi: 10.5194/se-6-583-2015.

PUTZKE, J., PEREIRA, A.B., PUTZKE, M.L., 1998. **Moss communities of Rip Point in northern Nelson Island, South Shetlands Islands, Antarctica**. *Pes. Ant. Bras.* 3, 103–115.

R CORE TEAM, 2018. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

SCHAEFER, C.E., FRANCELINO, M.R., SIMAS, F.N.B., ALBUQUERQUE FILHO, M.R. **Ecossistemas costeiros e Monitoramento Ambiental da Antártica Marítima (2004)**. NEPUT, Viçosa, 192p.

SCHMITZ, D., SCHAEFER, C.E.R.G., PUTZKE, J., FRANCELINO, M.R., FERRARI, F.R., CORREA, G.R., VILLA, P.M. 2020. **How does the pedoenvironmental gradient shape non-vascular species**



**assemblages and community structures in Maritime Antarctica?** Ecological Indicators 108: 105726. <https://doi.org/10.1016/j.ecolind.2019.105726>

SIMAS, F.N.B., et al. 2004. **Atributos químicos de criossolos ornitogênicos da Baía do Almirantado, Antártica.** In: SCHAEFER, C.E., FRANCELINO, M.R., SIMAS, F.N.B., ALBUQUERQUE FILHO, M.R. Ecosistemas costeiros e monitoramento ambiental da Antártica Marítima. NEPUT, Viçosa, pp41-46.

SIMAS, F.N.B., SCHAEFER, C.E.G.R., MELO, V.F., ALBUQUERQUE-FILHO, M.R., MICHEL, R.F.M., PEREIRA, V.V., GOMES, M.R.M., DA COSTA, L.M., 2007. **Ornithogenic Cryosols from Maritime Antarctica: phosphatization as a soil forming process.** Geoderma 138, 191–203.

SLOWIKOWSKI, K., SCHEP, A., HUGUES, s., et al., 2019. **Package ggrepel:** Automatically Position Non-Overlapping Text Labels with 'ggplot2'. R package version 0.8.1.

THOMAZINI, A., FRANCELINO, M.R., PEREIRA, A.B., SCHÜNEMANN, A.L., MENDONÇA, E.S., SCHAEFER, C.E.G.R., 2018. **The spatial variability structure of soil attributes using a detailed sampling grid in a typical periglacial area of Maritime Antarctica.** Environ. Earth Sci. 77, 637. <https://doi.org/10.1007/s12665-018-7818-3>.

VINCINI, L. 2005. **Análise Multivariada da teoria à prática.** Tese de doutorado. Universidade Federal de Santa Maria (UFSM). 215p.

WEI, T. 2017. **Package corrplot:** Visualization of a Correlation Matrix. R package version 0.8.4.

WICKHAM, H. & CHANG, W. 2019. **Package ggplot2:** Create elegant data visualisations using the grammar of graphics. R package version 3.2.1.



## DISTRIBUIÇÃO ESPACIAL DE FATORES AMBIENTAIS E ATRIBUTOS FLORESTAIS USANDO ROTINAS NO R

Data de aceite: 12/02/2020

### Alice Cristina Rodrigues

Universidade Federal de Viçosa - Departamento de Biologia Vegetal - Laboratório de Ecologia e Evolução de Plantas - LEEP, Viçosa, Minas Gerais, Brasil.

### Pedro Manuel Villa

Universidade Federal de Viçosa - Departamento de Engenharia Florestal- Laboratório de Restauração Florestal – LARF, Viçosa, Minas Gerais, Brasil.

### Andreza Viana Neri

Universidade Federal de Viçosa - Departamento de Biologia Vegetal -Laboratório de Ecologia e Evolução de Plantas - LEEP, Viçosa, Minas Gerais, Brasil.

**RESUMO:** Compreender os mecanismos subjacentes à contribuição relativa de diferentes fatores ambientais na montagem das comunidades vegetais é fundamental para se estabelecer estratégias de conservação e manejo. Para esse fim, a aplicação de análises estatísticas tem sido uma importante ferramenta para elucidar os padrões de distribuição dos fatores bióticos e abióticos que podem determinar a estrutura e diversidade das comunidades florestais. Com isso, o objetivo desse capítulo é demonstrar uma metodologia de análise e construção de gráficos de distribuição

de fatores ambientais e atributos florestais usando os pacotes estatísticos ‘fields’, ‘raster’ e ‘ggplot2’ no software R. Serão construídos passo a passo três tipos de gráficos, gráfico do tipo *grid*, gráfico tridimensional com contornos de nível e gráficos com distribuição das variáveis em *raster*. Apresentaremos ao longo desse capítulo toda a metodologia para a construção dos gráficos. Desde a obtenção das variáveis em campo, passando pela sistematização da planilha de dados, descrição de todos os comandos utilizados nos *scripts* e carregamento dos dados no R até o produto final das análises. Na maioria dos estudos que utilizaram essa ferramenta, essas análises foram utilizadas para determinar a formação de habitats relacionados à variação dos atributos florestais. Essa metodologia é interessante para estudos onde se espera demonstrar associações de espécies e atributos da vegetação com habitats ou gradientes de variáveis ambientais. Sendo também útil em análises de processo de distribuição de nichos e regras de montagem de comunidades.

**PALAVRAS-CHAVE:** Fatores bióticos e abióticos, elevação, classificação de habitats, estrutura florestal, gradiente ambiental, *Kriging*.

SPATIAL DISTRIBUTION OF  
ENVIRONMENTAL FACTORS AND FOREST

**ABSTRACT:** Understanding the mechanisms underlying the relative contribution of different environmental factors in the assembly communities is fundamental for establishing conservation and management strategies. To this end, the application of statistical analysis has been an important tool to elucidate the distribution patterns of biotic and abiotic factors that can determine the structure and diversity of forest communities. Thus, the purpose of this chapter is to demonstrate a methodology for analyzing and constructing a distribution graph of environmental factors and forest attributes using the 'fields', 'raster' and 'ggplot2' statistical packages in the R software. Three types of graphs will be constructed, grid graph, three-dimensional contoured graph, and raster variable distribution graphs. We will present throughout this chapter all the methodology for the construction of graphs. From obtaining the variables in the field, through the systematization of the dataset, description of all commands used in the scripts and loading of data in R to the final product of the analysis. In most studies using this tool, these analyses were used to determine the formation of habitats related to the variation of forest attributes. This methodology is interesting for studies where it is desired to demonstrate species associations and vegetation attributes with habitats or gradients of environmental variables. Also useful in niche formation process analysis and assembly community rules.

**KEYWORDS:** Biotic and abiotic factors, elevation, habitat classification, forest structure, environmental gradient, Kriging.

### 1 | INTRODUÇÃO

Conhecer a contribuição relativa dos diferentes fatores que podem determinar a estrutura e diversidade das comunidades florestais tem sido um tema relevante na ecologia contemporânea (Rozendaal et al. 2019). Recentemente o impacto desses fatores ambientais, como solo, topografia e clima tem ganhado importância para conhecer melhor as florestas tropicais (Yuan et al. 2016; Rozendaal et al. 2019). Por exemplo, os estoques de biomassa acima do solo variam amplamente entre as florestas devido aos efeitos diferenciais dos fatores abióticos (como topografia, fertilidade do solo e clima) e bióticos (como a biodiversidade e os atributos estruturais da floresta).

Contudo, para compreender os mecanismos subjacentes à contribuição relativa de diferentes fatores ambientais na montagem das comunidades vegetais, é fundamental a aplicação de análises estatísticas com métodos que nos permitam elucidar os padrões de distribuição dos fatores bióticos e abióticos em nossas áreas amostrais. Sendo esse conhecimento, imprescindível para se estabelecer estratégias de conservação e manejo, bem como prever as respostas das comunidades vegetais à variabilidade de fatores ambientais e a mudanças climáticas.

O objetivo desse capítulo é demonstrar uma metodologia de análise e construção de gráficos de distribuição de fatores ambientais e atributos florestais usando o software R. Construímos três tipos de gráficos, 1) tipo *grid*, 2) o tridimensional com contornos de nível e 3) com distribuição das variáveis em *raster*. Iniciaremos pela demonstração de como obter as variáveis em campo, passando pela sistematização da planilha de dados, descrição de todos os comandos utilizados nos *scripts* e carregamento dos dados no R até o produto final das análises.

## 2 | METODOLOGIA

### 2.1 Obtenção das variáveis para análises

Para o desenvolvimento dos gráficos será utilizado como modelo uma área amostral que possui diferentes condições topográficas e está situada em um fragmento em estágio avançado de regeneração secundária de Floresta Atlântica. A amostra constituída de uma parcela de 1 ha, subdividida em 100 parcelas de 10 x 10 m (parcelas contiguas), abrangendo um gradiente topográfico desde o vale até o platô. Para esse capítulo serão utilizados, uma variável topográfica (elevação), um parâmetro químico do solo (pH) e um atributo estrutural das árvores (altura média), de cada parcela para a construção dos gráficos. Para amostragem de solo e atributos estruturais da vegetação, sugerimos a utilização de padrões usuais de amostragem. Para o levantamento das variáveis topográficas nas áreas de interesse, sugere-se a utilização de estação total para áreas menores que 5 ha e de *softwares GIS* (*Geographic Information System*) para áreas maiores.

No presente estudo, foi utilizado uma estação total para a obtenção das variáveis topográficas. Visto que a área de interesse possuía 1 ha e perderíamos a acurácia dos dados utilizando softwares GIS. Para cada parcela amostrada foram medidas e calculadas três variáveis topográficas (elevação, declividade e convexidade). A elevação foi calculada usando o valor médio de elevação de cada um dos quatro cantos da parcela. A declividade ( $^{\circ}$ ) foi medida como o desvio angular médio da horizontal de cada um dos quatro planos triangulares, formado pela conexão de três de seus cantos. A convexidade foi determinada subtraindo a elevação do centro da parcela da elevação média das oito parcelas circundantes, seguindo os métodos propostos por Wang et al., (2016).

### 2.2 Organização da planilha para análise no R

Após a obtenção das variáveis em campo, ou a partir de bancos de dados, é necessário organizá-las de forma sistemática em uma planilha eletrônica Excel. No material suplementar é mostrado como as variáveis devem ser sistematizadas na

planilha (link).

## 2.3 Pacotes estatísticos e scripts

Para a construção dos gráficos, serão utilizados três pacotes: *'fields'* que é utilizado para ajustes de curvas, superfícies e funções com ênfase em dados espaciais, geostatística e estatística espacial (Nychka, et. al., 2019). O pacote *'raster'*, que faz manipulações, análises e modelagem de dados espaciais em *grids* (Hijmans, et. al., 2019) e *'ggplot2'*, um pacote do R para criar gráficos baseados em mapeamento de atributos estéticos de formas geométricas (Wickham, et. al., 2019).

### 2.3.1 Descrição do comando ou script

A seguir descreveremos todos os comandos utilizados no script, desde o carregamento dos dados no R até o produto final das análises.

O primeiro passo é indicar endereço do arquivo ou a pasta de trabalho onde se encontra os dados/planilhas a serem analisadas. Nessa pasta também serão salvos todos os gráficos gerados nas análises.

```
> setwd (“~/endereço do arquivo”)
```

O comando a seguir é utilizado para discriminar a planilha em formato txt que será objeto de análise.

```
> dados <- read.table (“planilha x.txt”, h=T)
```

A função *attach()* é utilizada como uma maneira de extrair partes do nosso conjunto dados de uma *dataframe* ou de tornar todas as variáveis de um conjunto de dados acessíveis. Vejamos como funciona essa função:

Se digitamos apenas *dados* na linha de comando a seguinte mensagem irá aparecer:

```
> dados
```

```
Error: object “dados” not found
```

Utilizando a função *attach*, e digitando *dados* novamente, os dados do conjunto de dados ficarão agora disponíveis. Atenção: O nome *dados* é um nome fictício, você pode chamar seu conjunto de dados de *data*, por exemplo.

```
> attach (data)
```

```
> data # agora todo o conjunto de dados está disponível
```

Se quisermos, por exemplo, que apenas os dados referentes a *elevação* fiquem disponíveis devemos digitar o seguinte comando:

```
> attach (elevação)
```

```
> elevação # agora os dados elevação estão disponíveis
```

Atenção: A função *attach* pode trazer problemas. Se a utilizarmos em dois *dataframes*, e essas possuírem variáveis com o mesmo nome, corre-se o risco de

usarmos por engano a variável errada. Sempre que utilizarmos a função *attach* é preciso “desatachar” o objeto imediatamente após o seu uso. Para isso utilizamos a função *detach()*.

```
> detach (dados)
dados # não está mais disponível
> detach (elevação)
elevação # não está mais disponível
```

```
> attach (dados)
#Após utilizá-los#
> detach (dados)
```

Sempre que quisermos consultar nosso conjunto de dados, as funções a seguir nos permitem visualizá-los, assim como as dimensões da matriz (função *dim*) e o nome e resumo das variáveis que estamos utilizando nas análises (funções *names* e *summary*).

```
> dados
> dim (dados)
> names (dados)
> summary (dados)
```

Antes de começar a análise propriamente dita dos dados, é necessário a instalação dos pacotes, podendo estes serem instalados com a seguinte função:

```
> install.packages (“nome do pacote”)
>install.packages(fields)
>install.packages(raster)
>install.packages(ggplot2)
```

Após instalar o pacote, é necessário “carregar” o pacote sempre que você abrir o R e for utilizá-lo. A função *library* é utilizada para “carregar” pacotes.

```
> library (fields)
> library (raster)
> library (ggplot2)
```

### 2.3.2 Resultados e comandos do estudo de caso

### 2.3.3 Construção gráfico grid com a função *image.plot*

Feito os passos anteriores é possível agora começar a construção dos gráficos de interesse. A primeira tarefa é escolher em qual formato o gráfico será salvo (*jpeg*, *png*, *tiff*, entre outros), nesse estudo de caso o formato escolhido foi *tiff*. A função *file* permite nomear o arquivo (gráfico) que será salvo na pasta de trabalho previamente escolhida no diretório ou *setwd*.

Os argumentos *res*, *width*, *height* e *compression* correspondem à resolução,



largura, altura e compressão do gráfico, respectivamente, e podem ser ajustadas conforme nosso interesse. A função *par* permite estabelecer o número de colunas e linhas para disposição de gráficos como uma única imagem composta. No nosso caso, foram escolhidas uma linha e uma coluna para a construção de um único gráfico de forma individual. Esta função *par* é muito útil quando se quer apresentar mais de um gráfico na mesma figura (ex., apresentar dois gráficos na mesma linha e em duas colunas), como faremos mais adiante nesse capítulo.

A função *image.plot* é utilizada para adicionar legenda no gráfico *grid*. Os argumentos *xlab* e *ylab* são utilizados para dar nomes aos eixos x e y respectivamente. Para colocar título no gráfico use o argumento *main*. Os argumentos *font.lab*, *font.main*, *cex.axis* e *cex.main* permitem estabelecer a fonte/tamanho dos eixos e legendas. O argumento *byrow* é utilizado para determinar se a matriz da legenda será preenchida por colunas (*byrow=F*) ou por linhas (*byrow=T*). Use a função *dev.off* para exportar a figura para a pasta de trabalho previamente selecionada em *setwd*.

```
> tiff (file = "nome_do_arquivo_a_ser_salvo.tiff", res = 300, width = 4400, height = 3100, compression = "lzw")
> par (mfrow=c (1,1))
> image.plot (0:10*10,0:10*10, matrix (as.numeric(elevação) ,10,10,byrow=T), xlab="", ylab="",font.lab=2,cex.axis=1.25,main="Elevação",font.main=4,cex.main=2)
> dev.off( )
```

Utilizamos o conjunto de dados do plot de 1 ha de Floresta Atlântica para a construção do gráfico *grid* (Figura 1A), utilizando dados referentes à elevação. O argumento *matrix(as.numeric())* nos permite escolher dentro da matriz o conjunto de dados que desejamos plotar no gráfico.

#### 2.3.4 Construção do gráfico tridimensional com contornos de nível utilizando a função *drape.plot*

Para a construção do gráfico 3D (figura 1B) segue-se os mesmos argumentos utilizados para o gráfico *grid*. Porém, nesse caso iremos utilizar a função *drape.plot* ao invés de *image.plot*. A Função *drape.plot* é utilizada para produzir gráficos do tipo *wireframe*, com as facetas sendo preenchidas por cores diferentes. Os argumentos *theta* e *phi* são utilizados para determinar a rotação dos ângulos/eixos x-y e z, respectivamente. No presente estudo de caso, utilizamos nos eixos x e y as dimensões da área de estudo e no eixo z a elevação da área de estudo. O argumento *zlab* é utilizado para nomear o eixo z. Os argumentos *shade* e *col* permitem fazer sombreamento e degrade de cores no gráfico.

```

> tiff (file = "nome_do_arquivo_a_ser_salvo.tiff", res = 300, width = 4400, height
= 3100, compression = "lzw")

> par (mfrow=c (1,1))

> drape.plot (1:10,1:10,matrix (as.numeric(elev), 10, 10, byrow=T), theta =1023,
phi = 50, expand = 0.64,shade = 1.7,col= terrain.colors (128),xlab="x=100 m", ylab=
"y=100 m", zlab= "z=elevação", font.lab=2, cex.lab=1.3, main= "Mapa de Topografia",
cex.main=1.6, font=2)

> dev.off()

```

*2.3.5 Construção do gráfico grid com a função image.plot e do gráfico tridimensional com contornos de nível utilizando a função drape.plot na mesma figura usando a função par.*

Para a construção de dois gráficos na mesma imagem (figura 1) será utilizado a função *par* especificando o número de colunas e linhas que desejamos na figura.

Por exemplo:

```

> par (mfrow=c (1,1)) # Figura com uma linha e uma coluna.

> par (mfrow=c (1,2)) # Figura com uma linha e duas colunas.

> par (mfrow=c (2,2)) # Figura com duas linhas e duas colunas.

```

No presente estudo, construiremos dois gráficos, o que corresponde ao argumento *par* (*mfrow=c (1,2)*) para obtermos uma figura composta por uma linha e duas colunas (Figura 1). Note que, ao construirmos os dois gráficos é possível que tenhamos que ajustar as dimensões dos mesmos para que se adequem à imagem gerada, para isso podemos ajustar valores dos argumentos *width* e *height*.

```

> tiff (file = "nome_do_arquivo_a_ser_salvo.tiff", res = 300, width = 4400, height
= 2400, compression = "lzw")

> par (mfrow=c (1,2))

> image.plot (0:10*10,0:10*10, matrix (as.numeric(elevação) ,10,10,byrow=T),
xlab="", ylab="",font.lab=2,cex.axis=1.25,main="Elevação",font.main=4,cex.main=2)

> drape.plot (1:10,1:10,matrix (as.numeric(elev), 10, 10, byrow=T), theta =1023,
phi = 50, expand = 0.64,shade = 1.7,col= terrain.colors (128),xlab="x=100 m", ylab=
"y=100 m", zlab= "z=elevação", font.lab=2, cex.lab=1.3, main= "Mapa de Topografia",
cex.main=1.6, font=2)

> dev.off( )

```

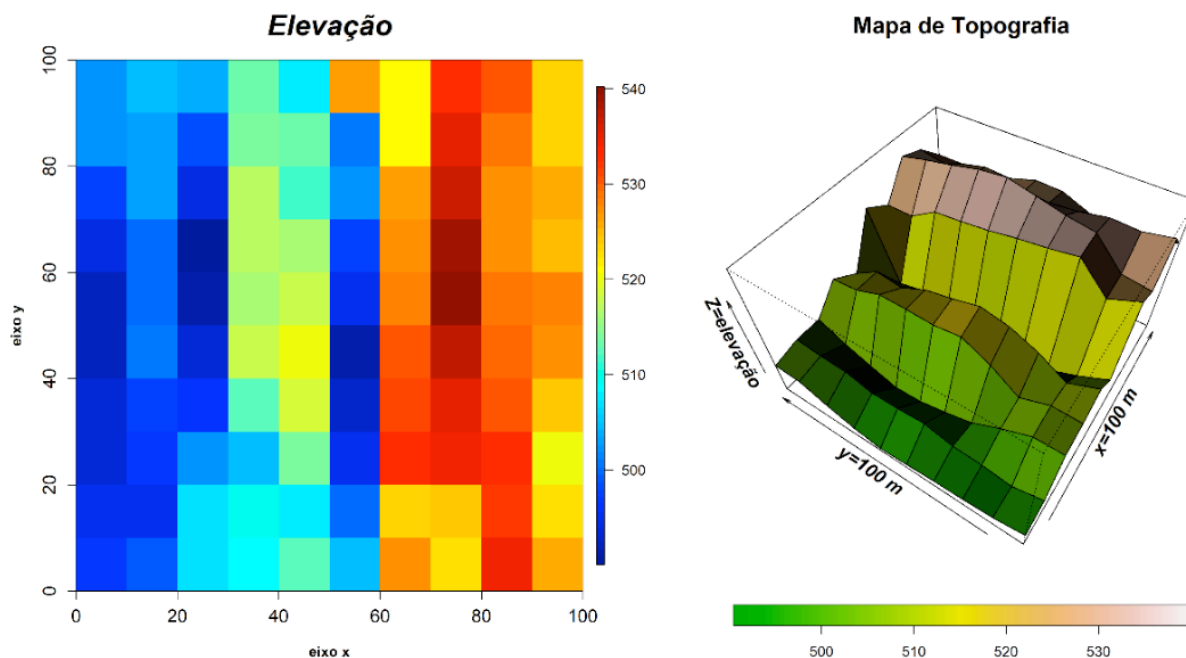


Figura 1. Gráfico grid e gráfico tridimensional com contornos de nível utilizando a mesma figura.

### 2.3.6 Construção gráfico grid com a função `image.plot` utilizando outras variáveis

É possível construir gráficos *grid* utilizando diferentes variáveis ambientais (topografia, solo e etc.) e atributos florestais (altura, densidade de madeira, diâmetro das árvores, área basal e etc.). Basta apenas selecionar as variáveis de interesse na barra de comandos. Por exemplo, se quiséssemos plotar a altura média das árvores em cada parcela ao invés de elevação, teríamos que usar os seguintes comandos:

```
> tiff (file = "altura média.tiff", res = 300, width = 4400, height = 3100, compression = "lzw")
> par (mfrow=c (1,1))
> image.plot (0:10*10,0:10*10, matrix (as.numeric(altura) ,10,10,byrow=T), xlab="", ylab="",font.lab=2,cex.axis=1.25,main="altura",font.main=4,cex.main=2)
> dev.off( )
```

### 2.3.7 Construção de gráficos raster com a função `geom_raster`

Suponhamos que possuímos um *dataframe* de dados que possui três colunas: x, y e z. Onde x e y são as respectivas coordenadas (no presente estudo o número de parcelas nas coordenadas x e y, respectivamente) e z um valor contínuo (elevação, pH, altura, etc.). Esse valor de z poderá ser plotado como uma imagem no gráfico em relação às coordenadas x e y. Essa imagem pode ser definida como um preenchimento em gradiente e pode ser plotada usando a função `geom_raster` () usando o pacote estatístico `ggplot2`. Vejamos alguns exemplos a seguir utilizando essa função:

Primeiramente segue-se os mesmos passos iniciais descritos anteriormente.

```
> setwd ("~/endereço do arquivo/")
> dados <- read.table ("planilha x.txt", h=T)
> attach (dados)
# Após utilizá-los #
> detach (dados)
```

Para a construção dos gráficos ‘raster’ será necessário a instalação do pacote ‘ggplot2’. Cabe ressaltar que para algumas análises será necessário também a instalação do pacote ‘raster’, pois, algumas funções ou argumentos podem estar conectados a este. Por isso, é recomendado a instalação de ambos os pacotes antes de se iniciar as análises.

```
> install.packages ("ggplot2/raster")
> library (raster)
> library (ggplot2)
```

A principal função do pacote ‘ggplot2’ é a função *ggplot ()* que permite construir gráficos peça por peça (passo a passo). Cada comando gráfico deve iniciar sempre com função *ggplot ()*. A função *aes ()* é usada para especificar eixos x e y (variáveis). Vejamos um exemplo:

Com os comandos básicos abaixo é possível criar um gráfico com os eixos x e y.

```
> ggplot (data, aes (x = x, y = y))
```

Se quisermos acrescentar um terceiro eixo (eixo z) no gráfico (Figura 2), sendo z um valor contínuo ou discreto, esse valor de z poderá ser plotado como uma imagem no gráfico em relação às coordenadas x e y utilizando os seguintes comandos:

```
> ggplot (data, aes (x = x, y = y, z = elevação, fill=elevação)) + geom_raster
(interpolate = T)
```

Onde, o argumento *fill* corresponde à variável de interesse que queremos plotar no preenchimento do gráfico com interpolação (*interpolate = T*). Para obtermos um gráfico sem interpolação usamos *interpolate = F*. Para o estudo de caso do presente capítulo utilizamos os dados de elevação.

Para fins de demonstração, iremos construir quatro gráficos do tipo ‘raster’. Utilizaremos como variáveis altura média das árvores em cada parcela, elevação e pH. Para diferenciar os gráficos, iremos utilizar os objetos *mod1*, *mod2*, *mod3* e *mod4*, como sendo gráficos 1 (Figura 2A), 2 (Figura 2B), 3 (Figura 3A) e 4 (Figura 3B), respectivamente. Nos comandos, o argumento *scale\_fill\_gradientn ()* permite escolher o gradiente de cores que desejamos para cada gráfico. Nos exemplos abaixo escolhemos diferentes gradientes de cores para exemplificar. O argumento *theme*, permite escolher a formatação do gráfico, nesse caso escolhemos o formato *gray*. Dentro da função *theme\_gray ()* pode-se ajustar automaticamente o tamanho

das letras das legendas, eixos e linhas usando *base\_size*. Um tamanho razoável para artigos científicos é entre 14 e 16. O argumento *labs*, é utilizado para dar título ao eixo x e y respectivamente. Utilizamos *aes* (*fill = elevação*), para construir a legenda de barra de cores e *scale\_x\_discrete* (*limits = c*) para estabelecermos a escala dos eixos x e y. No presente estudo estabelecemos uma escala de 0 a10, correspondente às 100 parcelas (10 x 10 m) da nossa área amostral (100 x 100 m).

```
> mod1 <- ggplot (data, aes (x = x, y = y, z = elevação, fill =elevação)) +  
geom_raster (interpolate = T) + scale_fill_gradientn(colours = rainbow(10)) +  
theme_gray (base_size = 14) + labs(title="", x="Número de parcelas", y = "Número de  
parcelas") + aes (fill = elevação) + scale_x_discrete(limits = c(0,1,2,3,4,5,6,7,8,9,10)) +  
scale_y_discrete(limits = c(0,1,2,3,4,5,6,7,8,9,10))  
> mod1  
  
> mod2 <- ggplot (data, aes (x = x, y = y, z = pH, fill =pH)) + geom_raster  
(interpolate = T) + scale_fill_gradientn(colours = rainbow(10)) + theme_gray (base_size  
= 14) + labs(title="", x="Número de parcelas", y = "Número de parcelas") + aes (fill =  
pH) + scale_x_discrete(limits = c(0,1,2,3,4,5,6,7,8,9,10)) + scale_y_discrete(limits =  
c(0,1,2,3,4,5,6,7,8,9,10))  
> mod2
```

Podemos construir também gráficos utilizando a função *geom\_contour*, essa função é muito útil, uma vez que o ggplot2 não pode desenhar superfícies 3D verdadeiras, mas usando a função *geom\_contour* podemos visualizar superfícies 3D em 2D. Para isso, os dados devem conter apenas uma linha para cada combinação exclusiva das variáveis mapeadas para o eixo x e y.

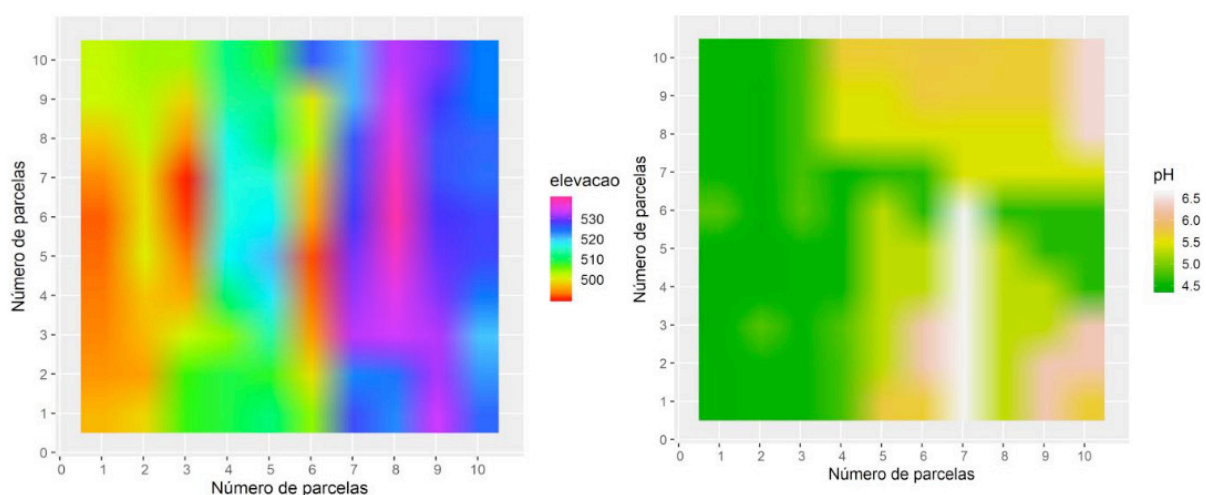


Figura 2. Gráficos raster com a função *geom\_raster* para elevação (A) e pH (B), utilizando diferentes escalas de cores.

O contorno tende a funcionar melhor quando x e y formam um grid uniformemente espaçado. Se seus dados não estiverem uniformemente espaçados, convém interpolar para formar um grid antes de visualizá-los (Wickham, et. al., 2019). Nos exemplos abaixo, construímos gráficos sem interpolação da elevação (Figura 3A)



e com interpolação da altura (figura 3B) utilizando diferentes escalas de cores, que podem ser ajustadas nos argumentos, conforme já descrito nesse capítulo.

```
> mod3 <- ggplot(data, aes(x=x, y=y, z=elevacao)) + theme_gray(base_size = 14) +
  geom_raster(aes(fill=elevacao), interpolate=F) + geom_contour (col='white', size=0.5)
+ labs(title="Sem interpolação", x="Número de parcelas", y = "Número de parcelas") +
  scale_x_discrete(limits=c(0,1,2,3,4,5,6,7,8,9,10))+scale_y_discrete(limits=
c(0,1,2,3,4,5,6,7,8,9,10))
> mod3
> mod4 <- ggplot(data, aes(x = x, y = y, z = altura))+theme_gray(base_size = 14)+
  geom_raster(aes(fill = altura), interpolate = T)+ scale_fill_gradient(low =
"white", high = "blue")+ geom_contour(colour = "black", binwidth = 1) + labs(fill =
"Altura", title = "Com interpolação", x = 'Número de parcelas', y = 'Número de parcelas')
+ scale_x_discrete (limits = c(0,1,2,3,4,5,6,7,8,9,10)) + scale_y_discrete(limits =
c(0,1,2,3,4,5,6,7,8,9,10))> mod4
```

Finalmente, as funções a seguir podem ser utilizadas para salvar os gráficos separadamente (função *ggsave*) ou de forma conjunta (função *grid.arrange*). Para essa última é necessário a instalação do pacote 'gridExtra'.

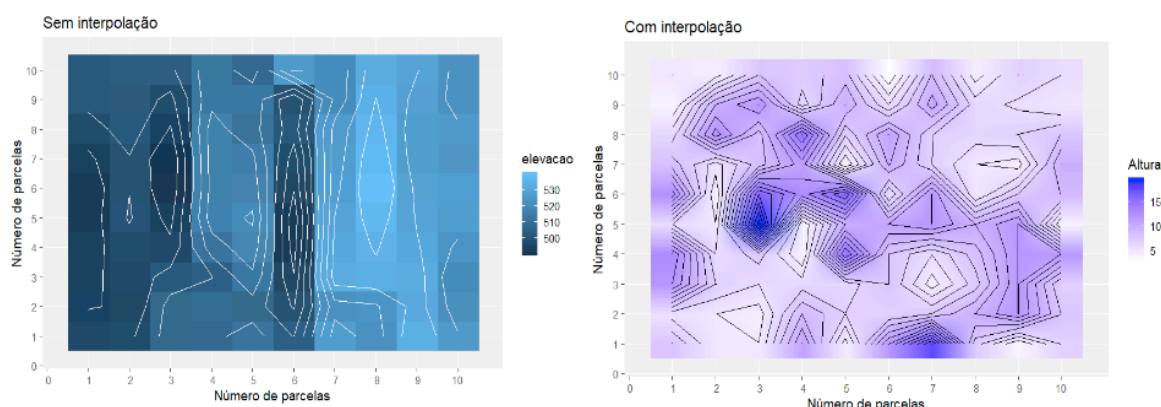


Figura 3. Gráficos raster com a função *geom\_raster* e *geom\_contour*, sem interpolação da elevação (A) e com interpolação da altura (B) utilizando diferentes escalas de cores.

```
> ggsave("nome_do_arquivo.png", width = 7, height = 6, dpi = 300)

> install.packages ("gridExtra")
> library(gridExtra)
> png("nome_do_arquivo.png", height=10, width=14, units="in", res=300)
grid.arrange(mod1, mod2, mod3, mod4, ncol = 2, nrow=2)

> dev.off()
```

### 3 | APLICAÇÃO DOS MÉTODOS EM ANÁLISES DE VEGETAÇÃO

Existem na literatura alguns trabalhos que utilizam essa metodologia de construção de gráficos descrita nesse capítulo (por exemplo: Liu, et al., 2014; Wang,



et. al., 2016; 2016; Rodrigues, et. al., 2019 a, b). Na maioria desses estudos, essas análises foram utilizadas para determinar a formação de habitats relacionados à distribuição espacial dos atributos florestais. Entre os habitats determinados, por exemplo, pela topografia, é provável que exista uma distribuição heterogênea de recursos, como água e nutrientes do solo, correlacionado com a existência de um gradiente edáfico. Essa heterogeneidade de distribuição dos fatores ambientais, na maioria desses estudos, foi relacionada diretamente com distribuição dos atributos da vegetação. Assim, esses gráficos constituem uma ferramenta interessante para demonstrar de forma criativa ao leitor a relação entre distribuição de variáveis ambientais e atributos da vegetação.

Analisando a figura 1A e 2A, observamos que há um gradiente de elevação, desde o baixio até o platô. A distribuição dessa variável está relacionada com a distribuição do pH, na mesma área de estudo (figura 2B), o qual pode correlacionar-se, por exemplo, com a altura das plantas (figura 3B). Liu, et. al., 2014 demonstrou em um estudo realizado no sul da China, que a composição de espécies, área foliar, altura máxima das árvores, a densidade de madeira e a massa das sementes relacionava-se com o gradiente de distribuição de variáveis ambientais tais com elevação, convexidade e declividade. Em um outro estudo, Wang et. al., 2016, encontrou uma relação entre variáveis topográficas e distribuição de espécies abundantes, raras, juvenis e adultos. Todos esses estudos usaram da ferramenta de construção de gráficos descrita nesse capítulo para mostrar de forma clara e objetiva esses achados.

Assim, essa metodologia de construção de gráficos pode ser bastante interessante para estudos onde se deseja demonstrar associações de espécies e atributos da vegetação com habitats ou gradientes de variáveis ambientais. Sendo também útil em análises de processo de formação de nicho e regras de montagem de comunidades.

## 4 | ACESSIBILIDADE AOS DADOS E COMANDOS NO R (SCRIPT)

[doi.org/10.13140/RG.2.2.33472.07686](https://doi.org/10.13140/RG.2.2.33472.07686)

### REFERÊNCIAS

Hijmans, R.J., et al. 2019. Raster: Geographic Analysis and Modeling with Raster Data. R Package Version 2.6-7. Available at <https://cran.r-project.org/web/packages/raster/index.html>.

Lan, G.Y., Hu, Y.H., Cao, M., Zhu, H., 2011. Topography related spatial distribution of

dominant tree species in a tropical seasonal rain forest in China. *Forest Ecology and Management*, 262:1507–1513.

- Liu, J., Yunhong, T., Slik, J.W.F. 2014. Topography related habitat associations of tree species traits, composition and diversity in a Chinese tropical forest. *Forest Ecology and Management* 330, 75–81.
- Meiners, S. J., Cadotte, M. W., Fridley, J. D., Pickett, S. T. A., & Walker, L. R. (2015). Is successional research nearing its climax? New approaches for understanding dynamic communities. *Functional Ecology*, 29(2), 154–164. <https://doi.org/10.1111/1365-2435.12391>.
- Nychka D, Furrer R, Paige J, Sain S (2019). ‘fields’ package: Tools for Spatial Data. R package version 2.4-5. <https://cran.r-project.org/web/packages/fields/fields.pdf>.
- R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rodrigues, A.C., Villa P.M., Neri, A.V., 2019. Fine-scale topography shape richness, community composition, stem and biomass hyperdominant species in Brazilian Atlantic forest. *Ecological Indicators* 102, 208–217, <https://doi.org/10.1016/j.ecolind.2019.02.033>.
- Rodrigues, A.C., Villa P.M., Ferreira Júnior, W., et al., 2019. Fine-scale habitat differentiation shapes the composition, structure and aboveground biomass but not species richness of a tropical Atlantic forest. *Journal Forestry Research*, <https://doi.org/10.1007/s11676-019-00994-x>.
- Rozendaal, D. M. A., Bongers, F., Aide, T. M., Alvarez-Dávila, E., Ascarrunz, N., Balvanera, P., Becknell, J. M., Bentos, T. V., Brancalion, P. H. S, Cabral, G. A. L., Calvo-Rodriguez, S., Chave, J., et al. (2019). Biodiversity recovery of Neotropical secondary forests. *Science Advance*, 5:3114.
- Wang, Q., PUNCHI-MANAGE R., Lu, Z., Franklin, S.B., Wang, Z., Li, Y., Chi, X., Bao, D., Guo, Y., Lu, J., Xu, Y., Qiao, X., Jiang, M. 2016 Effects of topography on structuring species assemblages in a subtropical forest. *Journal of Plant Ecology* 10 (3): 440-449.
- Wickham, H. 2019. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.

## SPATIAL RELATIONSHIP BETWEEN SOIL AND PHYTOSOCIOLOGICAL INDICATORS OF ECOLOGICAL RESTORATION IN AN ATLANTIC FOREST SITE

*Data de aceite: 12/02/2020*

*Data de submissão: 21/11/2019*

### **Camila Santos da Silva**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/5813811233254259

### **Marcos Gervasio Pereira**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/3657759682534978

### **Rafael Coll Delgado**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/1178948690201659

### **Emanuel José Gomes de Araújo**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/3124188823262093

### **Cristiane Figueira da Silva**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/5734423515627169

### **Daniel Costa de Carvalho**

Universidade de Brasília, Brasília – DF  
lattes.cnpq.br/6394423734483181

### **Shirlei Almeida Assunção**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/3592454473382052

### **Israel Oliveira Ramalho**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/2994569158207067

### **Deyvid Diego Carvalho Maranhão**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/7821807823585233

### **Ariovaldo Machado Fonseca Junior**

Universidade Federal Rural do Rio de Janeiro,  
Seropédica – RJ  
lattes.cnpq.br/3170882049619127

**ABSTRACT:** The objective of this study was to assess spatial relationships between soil and phytosociological indicators in an Atlantic Forest site under ecological restoration. Soil samples (0–10 cm depth) were collected from 49 points within the study area. Total organic carbon (TOC), particulate organic carbon (POC), mineral-associated organic carbon (MOC), permanganate oxidizable carbon (POXC), total nitrogen, glomalin-related soil protein easily extractable fraction (GRSP-EF), and arbuscular mycorrhizal fungal spore abundance were determined. The structure and diversity of tree and regenerating individuals were also assessed. Semivariogram data were subjected to ordinary kriging, and prediction maps were generated. All variables were spatial dependent. A high spatial correlation was observed between

GRSP-EF and density of tree and regenerating individuals. Glomalin has important functions in the soil, favoring the development of microorganisms and plants. TOC, POC MOC, and POXC were spatially correlated with structure and diversity parameters of tree individuals. These relationships were likely due to the presence of leaf litter, which contributes to soil organic matter content. Pearson correlation analysis showed that related variables were highly correlated. However, low or insignificant correlations were found among vegetation structure and soil attributes. Overall, phytosociological indicators showed little correlation with soil indicators.

**KEYWORDS:** soil attributes, phytosociological characterization, geostatistics, forest planting, forest restoration.

## RELAÇÃO ESPACIAL ENTRE INDICADORES EDÁFICOS E FITOSSOCIOLÓGICOS EM UMA RESTAURAÇÃO ECOLÓGICA NA FLORESTA ATLÂNTICA

**RESUMO:** O objetivo do presente estudo foi relacionar espacialmente indicadores do solo e fitossociológicos em uma área em processo de restauração ecológica, em uma floresta da Mata Atlântica. Foram coletadas amostras de solo (0-10 cm) distribuídas em 49 pontos na área de estudo. Em cada amostra foram determinados: o carbono orgânico total (COT), carbono orgânico particulado (COp) e associado aos minerais (COam), carbono oxidável permanganato (POXC), nitrogênio total, proteína do solo relacionada à glomalina – facilmente extraível (PSRG-FE) e abundância de esporos de fungos micorrízicos arbusculares. Nos componentes arbóreo e regenerante foram determinados os parâmetros de estrutura e diversidade. O semivariograma foi ajustado para cada variável, posteriormente foi realizada a Krigagem Ordinária. Todas as variáveis apresentaram dependência espacial. De acordo com a espacialização, a PSRG-FE apresentou correlação espacial com a densidade de indivíduos arbóreos e regenerantes. Essa relação pode ter ocorrido porque a glomalina possui diferentes funções no solo, favorecendo o desenvolvimento dos organismos e plantas. O COT, COp COam e POXC apresentaram correlações espaciais com os parâmetros de estrutura e diversidade do componente arbóreo. Esse padrão, provavelmente, foi encontrado devido à deposição de resíduos vegetais, contribuindo com a adição de matéria orgânica no solo. Em relação à correlação de Pearson, as variáveis de mesma natureza apresentaram maiores correlações entre si, enquanto a estrutura da vegetação apresentou correlações baixas e insignificantes com os indicadores edáficos. Em geral, os indicadores relacionados aos atributos do solo apresentaram pouca correlação com os parâmetros fitossociológicos dos componentes arbóreos e regenerantes.

**PALAVRAS-CHAVE:** atributos do solo, caracterização fitossociológica, geoestatística, plantio florestal, restauração florestal.

## 1 | INTRODUCTION

Centuries of deforestation have reduced the Atlantic Forest to about 8.5% of its original cover (Fundação SOS Mata Atlântica, 2016). This situation has prompted an increasing interest in the preservation of forest landscapes (Miranda et al., 2007), evidenced by the growing number of ecological restoration initiatives seen in Brazil.

Ecological restoration can be defined as the process of assisting in the recovery of an ecosystem that has been disturbed, damaged, or degraded. The aim may vary from protecting biodiversity to restoring ecosystem services (SER, 2004). A critical step in ecological restoration is the evaluation and monitoring of the forest site (Brancalion et al., 2015). Different parameters may need to be examined to determine whether the site is responding positively to the adopted strategies. The most commonly used indicators are related to plant community structure and composition (Brancalion et al., 2015), but chemical, physical, and biological attributes of soil can also be used. If the results suggest low long-term sustainability, the restoration project may need to be revised (Brancalion et al., 2015).

Biological parameters often show spatial variability. However, most restoration studies disregard this information (Dormann, 2007). Structural attributes of tree individuals, such as height, basal area, biomass, carbon content, and species diversity, may show spatial dependence according to microclimatic conditions, species characteristics, and biotic factors (Rode et al., 2010).

The high spatial variability of variables has been a subject of geostatistical investigations (Yun-Qiang et al., 2009; Silva et al., 2017). In geostatistics, interpolation methods are used to predict variables for unsampled locations (Landim, 2006). This study aimed to investigate spatial relationships between soil and phytosociological indicators in an Atlantic Forest site under restoration in Brazil. We explored the following research question: How do soil attributes influence the spatial variations of forest structure? We hypothesized that plant species distribution depends on terrain relief, soil quality, and nutrient availability, that vegetation influences soil attributes, and that soil and phytosociological parameters are interdependent and spatially correlated.

## 2 | MATERIAL AND METHODS

The study area occupies 14,902 m<sup>2</sup> in the municipality of Magé, Rio de Janeiro, Brazil, under coordinates 43°01'W 22°35'S. Classified as Af in the Köppen system, the climate of the region is hot and humid, with no dry season. The average annual temperature is 21.9 °C and the average annual rainfall 2,050 mm. The soil is Gleysol

and Latosol (Santos et al., 2013), corresponding to Entisol and Oxisol, respectively, in the US Soil Taxonomy system (Soil Survey Staff, 2014). Dense ombrophilous forest is the primary vegetation (Fundação SOS Mata Atlântica/INPE, 2013), but, prior to restoration, the area was mostly covered by *Brachiaria* sp. and *Imperata brasiliensis* Trin.

Tree species native to the Atlantic Forest and an exotic tree species, *Khaya ivorensis* A. Chev. (African mahogany), were planted in 2009, with a spacing of 2.5 × 2.5 m. Seedlings received a starter fertilization of 150 g/hole NPK (06-30-06), 300 g/hole limestone, and 3 L/hole composted cattle manure.

Plot corners were georeferenced using mono- and dual-frequency global positioning system (GPS) receivers. Then, 49 sampling points, 20 m apart from one another, were systematically distributed across the area using ArcGIS 10.5. Sampling point coordinates were transferred to a GPS receiver, and soil samples were collected from the 0–10 cm depth.

Samples were evaluated for total organic carbon (TOC) content (Donagemma et al., 2011), particle size fractionation of soil organic matter (Cambardella and Elliot, 1992), permanganate oxidizable carbon (POXC) content (Weil et al., 2003), and total nitrogen content (Tedesco et al., 1995). The following biological parameters were also analyzed: arbuscular mycorrhizal fungal spore abundance (SA) (Gerdermann and Nicolson, 1963) and glomalin-related soil protein easily extractable fraction (GRSP-EF) (Wright and Upadhyaya, 1998; Rillig, 2004).

Plants were sampled from 6 m radius circular plots (113.10 m<sup>2</sup>) centered at each sampling point. All arboreal and regenerating plants were measured and identified. Tree individuals with a circumference at breast height equal to or greater than 15.7 cm (diameter at breast height, DAB ≥ 5 cm) were analyzed. Plants that did not meet this criterion but had a height greater than 60 cm were considered regenerating individuals, and only the number of individuals of each species was recorded. Circumference and height measurements were taken with a metric tape and hypsometer, respectively.

In order to characterize the tree stratum, we calculated the following dendrometric and phytosociological indicators: height (m), basal area (m<sup>2</sup> ha<sup>-1</sup>), richness, density (ind ha<sup>-1</sup>), Shannon–Wiener diversity index ( $H'$ ), and number of individuals belonging to the family Fabaceae—the most important family for regeneration of the studied area. The following indicators were estimated for regenerating individuals: richness, density,  $H'$ , and number of Fabaceae individuals.

Dendrometric and phytosociological indicators were subjected to an experimental semivariogram analysis using R software 3.4.0 (R Core Team, 2017). The nugget effect, sill, and range were calculated, and spherical, exponential, and



Gaussian models were fitted by the maximum likelihood method using the geoR package in R (Ribeiro Júnior and Diggle, 2001). The used routine for geostatistical analysis we used R can be accessed and downloaded in the link: [doi.org/10.13140/RG.2.2.31036.80005](https://doi.org/10.13140/RG.2.2.31036.80005). The best-fitting semivariogram model was selected according to the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the coefficient of determination ( $r^2$ ), Willmott's concordance index (d), and the degree of spatial dependence (DSD). DSD was classified as follows: strong spatial dependence ( $DSD \leq 25\%$ ), moderate spatial dependence ( $25\% \leq DSD \leq 75\%$ ), and weak spatial dependence ( $DSD \geq 75\%$ ). A DSD of 100% indicates a pure nugget effect. Dendrometric and phytosociological parameters for unsampled locations were predicted by the ordinary kriging method using ArcGIS 10.5 software. Kriging maps were interpolated two by two, and a Pearson correlation matrix was constructed. This procedure allowed to assess the similarity between maps. Correlation ( $r$ ) was classified as very high ( $r \geq 0.90$ ), high ( $0.90 > r \geq 0.70$ ), moderate ( $0.70 > r \geq 0.50$ ), low ( $0.50 > r \geq 0.30$ ), or insignificant ( $r < 0.30$ ) (Mukaka, 2012).

### 3 | RESULTS AND DISCUSSION

Spatial dependence was observed for all variables (Table 1). Grego and Vieira (2005) found that physical and hydraulic properties of surface soil (0–25 cm depth) were spatial dependent, in agreement with the results of the current study. The spherical model, followed by the Gaussian model, showed the best fit to semivariograms. The exponential model had a poor fit to the data.

Parameter	Soil attributes						
	TOC	POC	MOC	POXC	N	GRSP-EF	SA
	-----g kg <sup>-1</sup> -----					mg g <sup>-1</sup>	spores 50 cm <sup>-3</sup> soil
AIC	333.59	212.36	317.47	-57.52	11.40	141.89	677.81
BIC	341.15	219.85	324.96	-50.48	18.96	149.45	685.38
$r^2$	0.09	0.14	0.00	0.00	0.00	0.21	0.18
d	0.57	0.54	0.38	0.34	0.46	0.65	0.64
C <sub>0</sub>	33.33	1.62	33.27	0.01	0.04	0.00	29855.73
C <sub>0</sub> +C	47.94	4.55	37.73	0.01	0.06	1.00	86103.26
a	67.58	41.84	39.69	41.84	39.12	33.39	45.75
DSD (%)	69.52	35.57	88.18	81.25	69.33	0.00	34.67
Model	Sph	Sph	Gaus	Sph	Sph	Sph	Gaus

Tree individuals						
	h (m)	G (m <sup>2</sup> ha <sup>-1</sup> )	S	D (ind ha <sup>-1</sup> )	H'	Fabaceae
AIC	195.77	283.33	214.12	636.14	93.31	198.74
BIC	203.09	290,37	221.44	643.45	100.63	206.31
r <sup>2</sup>	0.06	0.02	0.02	0.26	0.04	0.30
d	0.49	0.53	0.44	0.67	0.45	0.66
C <sub>0</sub>	1.73	28.57	4.78	24090.40	0.34	1.57
C <sub>0</sub> + C	3.62	36.65	5.34	58618.40	0.39	3.59
a	37.38	71.79	67.09	58.00	59.28	76.20
DSD (%)	47.73	77.96	89.52	41.10	87.78	43.71
Model	Sph	Sph	Gaus	Sph	Gaus	Sph

Regenerating individuals				
	S	D (ind ha <sup>-1</sup> )	H'	Fabaceae
AIC	140.67	563.49	30.61	227.38
BIC	147.00	569.82	36.95	234.94
r <sup>2</sup>	0.00	0.28	0.02	0.01
d	0.29	0.70	0.42	0.50
C <sub>0</sub>	1.74	0.00	0.08	3.92
C <sub>0</sub> + C	2.37	344343.01	0.11	6.05
a	43.58	39.74	51.88	59.42
DSD (%)	73.32	0.00	72.00	64.82
Model	Sph	Sph	Sph	Sph

Table 1. Semivariogram model parameters for soil and phytosociological variables of an Atlantic Forest site undergoing ecological restoration.

TOC, total organic carbon; POC, particulate organic carbon; MOC, mineral-associated organic carbon; POXC, permanganate oxidizable carbon; N, nitrogen; GRSP-EF, glomalin-related soil protein easily extractable fraction; SA, arbuscular mycorrhizal fungal spore abundance; h, height; G, basal area; S, richness; D, density; H', Shannon-Wiener diversity index; AIC, Akaike information criteria; BIC, Bayesian information criteria; C<sub>0</sub>, nugget effect; C<sub>0</sub> + C, sill; a, range; DSD, degree of spatial dependence; Sph, spherical; Gaus, Gaussian.

Our results corroborate previous reports. Yun-Qiang et al. (2009) used statistics and geostatistics to characterize and compare the spatial distribution of soil organic carbon and related factors in the Liudaogou basin, northern China. The authors found that the spherical model was best suited to describe the spatial dependence of soil organic carbon. Meng et al. (2009), in estimating the basal area of pines in Georgia using geostatistics, observed that the spherical model provided the best fit to data.

Most variables had moderate spatial dependence, but GRSP-EF and species density showed strong spatial dependence. Range varied from 33.39 (GRSP-EF) to 67.58 (TOC) for soil attributes, from 37.38 (h) to 76.20 (Fabaceae) for tree individuals, and from 39.74 (density) to 59.42 (Fabaceae) for regenerating individuals. Range represents the distance limit beyond which samples are no longer correlated. High

range indicates that sample points are influenced by other sample points located far apart (Isaaks and Srivastava, 1989). Thus, Fabaceae tree individuals correlate with each other even when at great distances apart.

Semivariogram parameters were used for ordinary kriging to generate prediction maps (Figures 1, 2, and 3). The distribution patterns in the northern region of the study site may be related to soil mineral content (Figure 1A–D). Carbon content is higher in soils with higher clay content. Soil organic matter interacts with the mineral fraction, resulting in the formation of organo-mineral complexes (Signor et al., 2014).

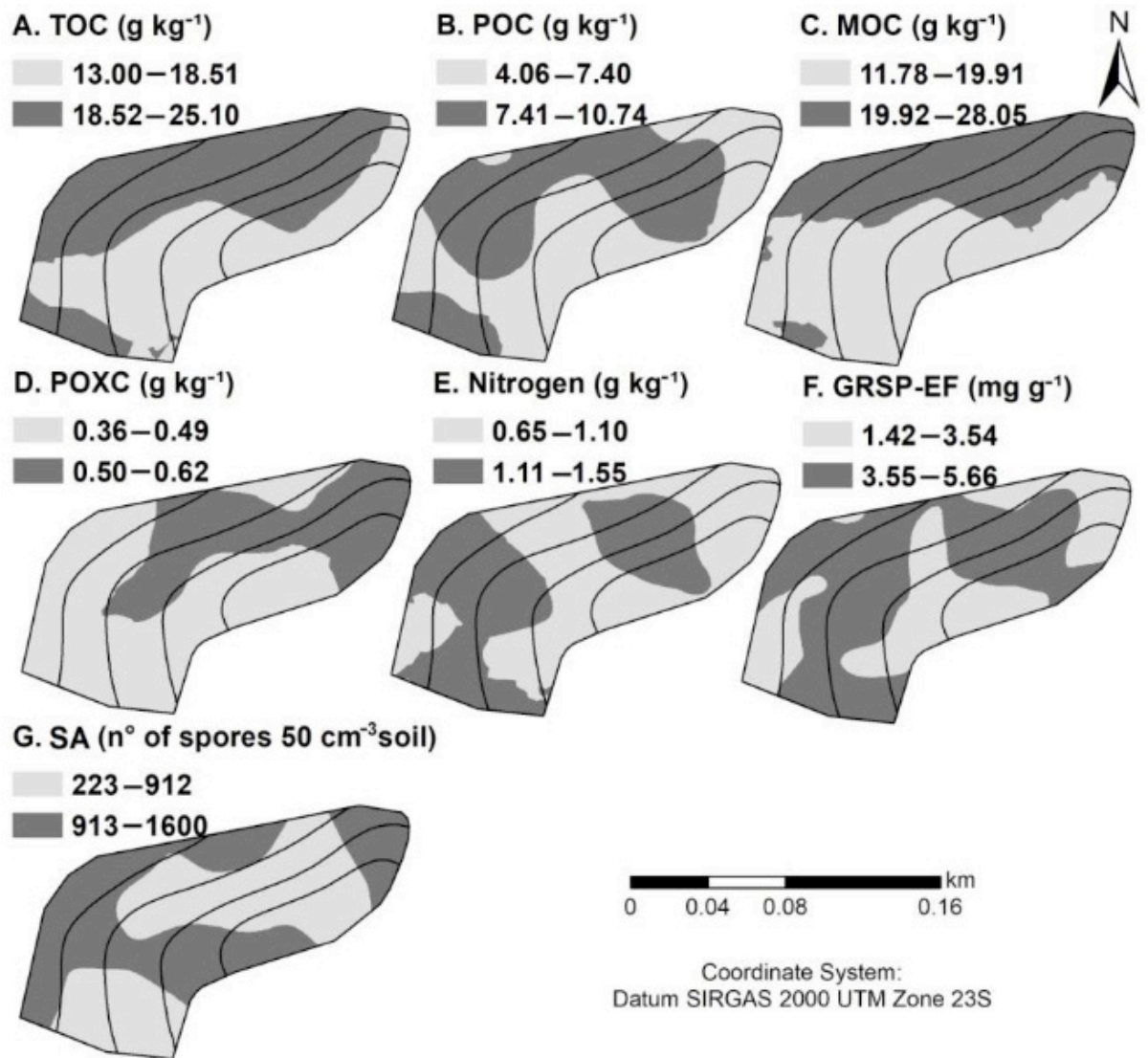


Figure 1. Spatial prediction maps of (A) total organic carbon (TOC), (B) particulate organic carbon (POC), (C) mineral-associated organic carbon (MOC), (D) permanganate oxidizable carbon (POXC), (E) nitrogen, (F) glomalin-related soil protein easily extractable fraction (GRSP-EF), and (G) arbuscular mycorrhizal fungal spore abundance (SA) in soil from an Atlantic Forest site under restoration. Contour lines indicate terrain elevation.

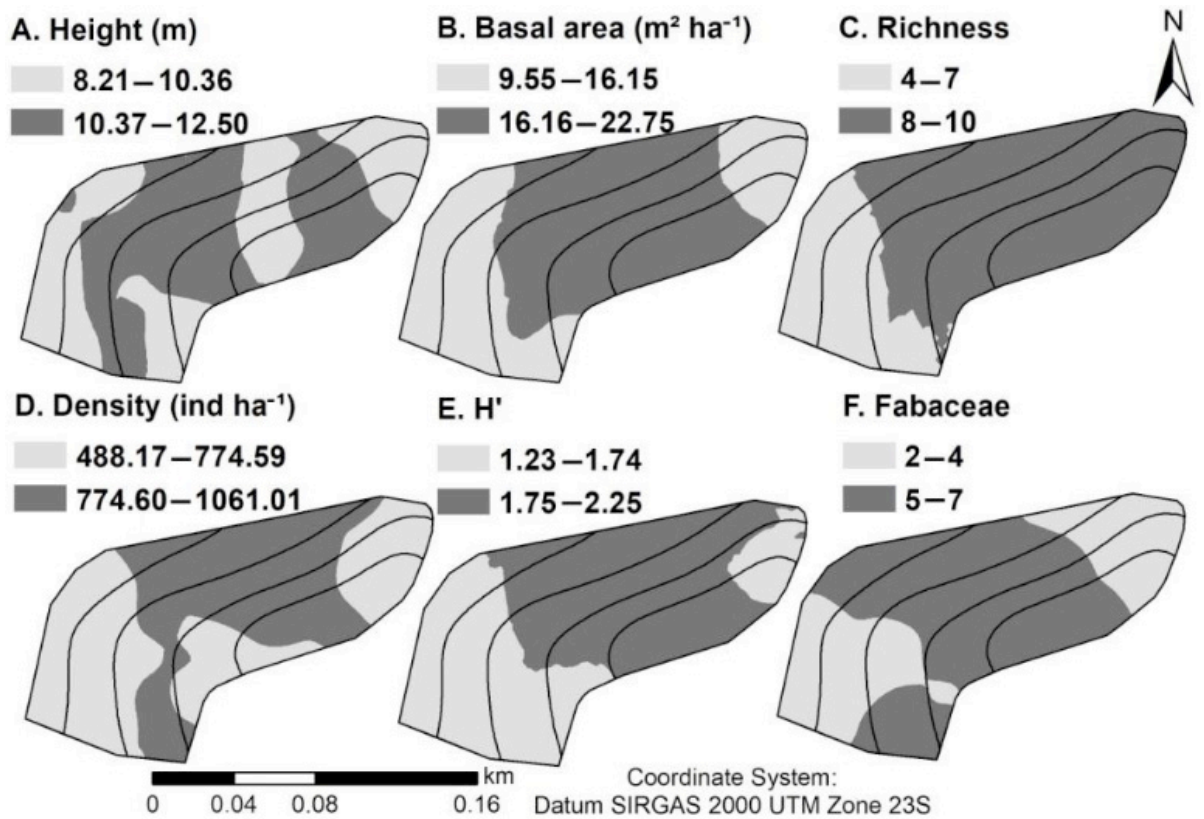


Figure 2. Spatial prediction maps of (A) height, (B) basal area, (C) richness, (D) density, (E) diversity ( $H'$ ), and (F) number of Fabaceae tree individuals growing in an Atlantic forest site under restoration. Contour lines indicate terrain elevation.

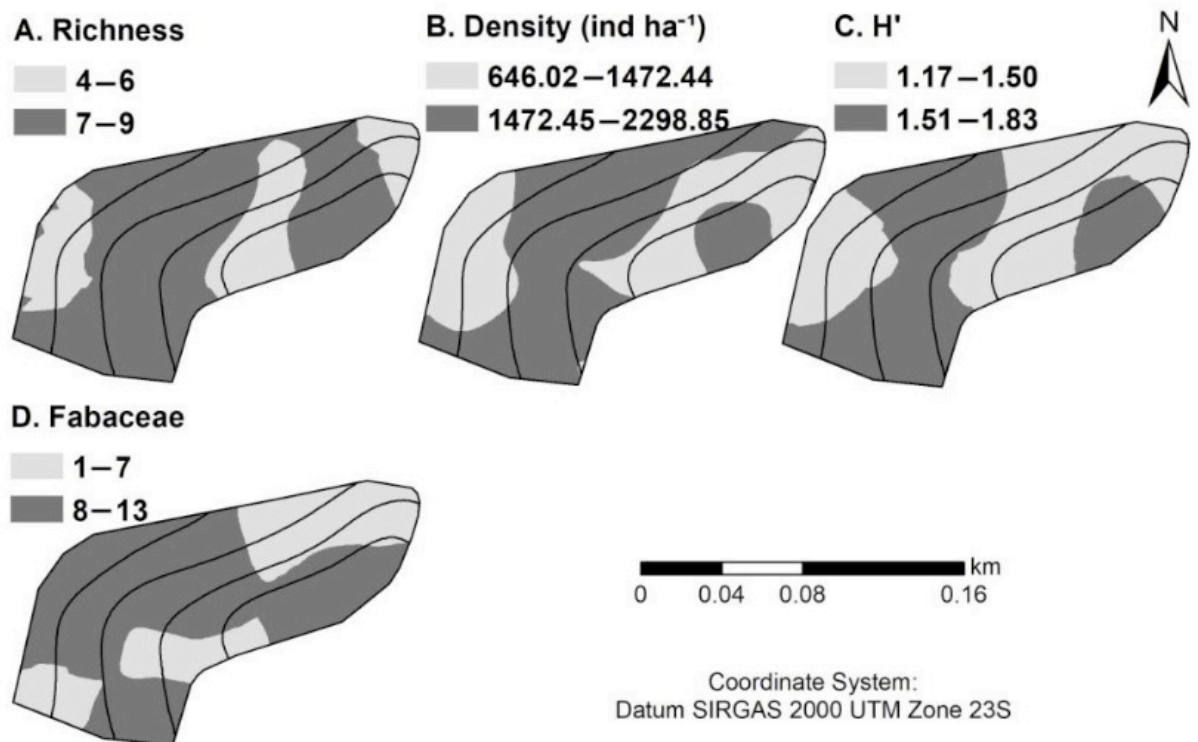


Figure 3. Spatial prediction maps of (A) richness, (B) density, (C) diversity ( $H'$ ), and (D) number of Fabaceae regenerating individuals in an Atlantic Forest site under restoration. Contour lines indicate terrain elevation.

Soil samples had higher levels of MOC (Figure 1C) than POC (Figure 1B).

According to Bayer et al. (2004), 80% of the TOC is present in the MOC fraction, in agreement with the high correlation found between TOC and MOC in the present study ( $r = 0.93$ ). MOC interacts strongly with soil minerals, being easily accumulated in the soil. Similar results were reported by Guimarães et al. (2012). The authors found that Latosol under Atlantic Forest had higher levels of MOC ( $9 \text{ g kg}^{-1}$ ) than POC ( $4 \text{ g kg}^{-1}$ ) at the 0–10 cm depth. Leaf litter probably contributed to nutrient cycling and the increase in organic matter content, favoring accumulation of carbon in the MOC fraction.

POXC and SA (Figure 1D and G, respectively) were inversely correlated ( $r = -0.30$ ). Labile carbon consists of organic compounds that are easily mineralized by soil microfauna (Schimiguel et al., 2014). Therefore, areas with higher SA are those with the lowest oxidizable carbon content. GRSP-EF and SA, on the other hand, were positively correlated. Glomalin is a glycoprotein produced by mycorrhizal fungi; its levels in soil increase with the decomposition of spores and hyphae (Driver et al., 2005).

Similar distribution patterns were observed for basal area, richness, density, and  $H'$  (Figure 2), probably because these parameters are related. Areas with a high density of individuals typically have a greater presence of trees (basal area), richness, and, consequently, diversity. The low number of Fabaceae individuals in southwestern and eastern areas is due to the occurrence of *Brachiaria* sp. and bamboo (Figure 2). These invasive species prevent the establishment of native species at early stages of succession (Correia and Martins, 2015).

The density of tree individuals (Figure 2D) was lower than that of regenerating individuals (Figure 3B). Richness and  $H'$  were higher among tree individuals (Figure 2C and E) than among regenerating individuals (Figure 3A and C). Durigan et al. (2000) performed a phytosociological study in three vertical strata, upper stratum ( $\text{DAB} \geq 5 \text{ cm}$ ), intermediate stratum ( $1 \text{ cm} \leq \text{DAB} < 5 \text{ cm}$ ), and lower stratum ( $\text{DAB} < 1 \text{ cm}$ , minimum height of 10 cm), at the Caetetus Ecological Station, São Paulo, Brazil, and found a higher density of individuals in the lower ( $28,875 \text{ ind ha}^{-1}$ ) and intermediate ( $3,525 \text{ ind ha}^{-1}$ ) strata than in the upper stratum ( $1,080 \text{ ind ha}^{-1}$ ).  $H'$  and basal area, however, were greater in the upper stratum ( $2.41$  and  $31 \text{ m}^2 \text{ ha}^{-1}$ , respectively) than in the intermediate ( $1.83$  and  $2 \text{ m}^2 \text{ ha}^{-1}$ , respectively) and lower strata ( $2.01 \text{ m}^2 \text{ ha}^{-1}$ ).

Figures 2 and 3 show that areas with the largest numbers of Fabaceae individuals had the highest values of richness, density, and  $H'$ , indicating that most individuals were of this family. Roppa et al. (2012) assessed the richness, diversity, and composition of the woody flora of an Atlantic Forest site under spontaneous regeneration and found a high occurrence of Fabaceae individuals, about 20% of the sampled species.



The prediction maps of phytosociological variables for regenerating individuals had similar patterns (Figure 3), corroborating the high correlations found between the variables. This result indicates that a single phytosociological variable can be used to assess the progress of restoration, such as richness, which is more easily determined, optimizing time and labor (Brançalion et al., 2015).

Areas with higher soil GRSP-EF (Figure 1F) had higher densities of tree (Figure 2D) and regenerating individuals (Figure 3B). These results demonstrate the close relationship between mycorrhizal fungi and plants. Glomalin contributes to soil quality, favoring plant development (Al-Kaisi et al., 2014; Qin et al., 2017). TOC, POC, MOC, and POXC distributions were correlated with those of basal area, richness, density, and  $H'$  of tree individuals, which is likely due to the high amounts of litterfall and soil cover (Bayer et al., 2000).

TOC showed a very high and positive correlation with MOC ( $r = 0.93$ ; Table 2). This result shows that soil carbon occurs mainly in a recalcitrant form, indicating greater stability in the soil. GRSP-EF was highly correlated with nitrogen content ( $r = 0.80$ ), as indicated by the similarity between their spatial distributions (Figure 1E and F). The correlation between GRSP-EF and POXC was moderate ( $r = 0.68$ ). According to Fokom et al. (2012), glomalin is an important mode of storage of carbon and nitrogen in the soil; studies have shown that glomalin is positively correlated with carbon and nitrogen. Richness, density, and  $H'$  of tree individuals were highly to very highly correlated, as also observed for regenerating individuals. The number of Fabaceae tree and regenerating individuals had a moderate to low correlation with richness, density, and  $H'$ . This is probably because most tree species (40.21%) and regenerating individuals (21.88%) belonged to this family.

	Correlation coefficient																
	Soil attributes					Tree individuals							Regenerating individuals				
	TOC	POC	MOC	POXC	GRSP-EF	SA	N	h	G	S	D	$H'$	Fabaceae	S'	D'	$H''$	
POC	0.30																
MOC	0.93	-0.08															
POXC	0.27	0.56	0.07														
GRSP-EF	0.49	0.44	0.34	0.68													
SA	-0.16	-0.22	-0.08	-0.30	-0.27												
N	0.46	0.53	0.27	0.59	0.80	-0.22											
h	-0.01	0.31	-0.13	0.17	0.16	-0.47	0.14										
G	-0.19	0.47	-0.38	0.16	-0.01	-0.22	0.10	0.41									
S	0.14	0.29	0.04	0.00	0.05	-0.38	0.03	0.41	0.49								
D	0.19	0.29	0.09	0.02	0.18	-0.35	0.15	0.39	0.43	0.90							
$H'$	0.16	0.28	0.06	0.00	0.08	-0.38	0.05	0.48	0.48	0.96	0.85						
Fabaceae	0.12	0.40	-0.03	0.08	0.33	-0.22	0.32	0.36	0.30	0.47	0.57	0.47					
S'	0.00	0.12	-0.04	-0.09	-0.12	-0.26	-0.14	0.22	0.39	0.24	0.21	0.25	0.13				
D'	0.03	0.14	-0.03	-0.13	-0.14	-0.16	-0.21	0.22	0.48	0.35	0.42	0.33	0.18	0.78			
$H''$	0.04	0.00	0.04	-0.12	-0.11	-0.26	-0.16	0.14	0.23	0.21	0.20	0.22	0.15	0.96	0.72		
Fabaceae'	-0.22	0.33	-0.35	0.03	-0.03	-0.17	0.04	0.14	0.62	0.17	0.11	0.15	0.34	0.58	0.55	0.50	

Table 2. Pearson correlation matrix for soil and phytosociological indicators of ecological restoration in an Atlantic Forest site.

TOC, total organic carbon; POC, particulate organic carbon; MOC, mineral-associated organic carbon; POXC, permanganate oxidizable carbon; N, nitrogen; GRSP-EF, glomalin-related soil protein easily extractable fraction;



SA, arbuscular mycorrhizal fungal spore abundance; h, height; G, basal area; S, richness; D, density; H', Shannon–Wiener diversity index. \*Results for regenerating individuals.

## 4 | CONCLUSION

The results of this study show that correlations between soil and phytosociological variables are high within each group but low or insignificant between groups. GRSP-EF was correlated with density of tree and regenerating individuals. TOC, POC, MOC, and POXC had spatial correlations with basal area, richness, density, and H' of tree individuals. The fact that phytosociological variables were highly correlated suggests that a single variable can be used to assess the progress of forest restoration.

Overall, soil and phytosociological indicators were not spatially correlated. Further studies should investigate the relationship between soil and phytosociological indicators in other areas for the monitoring of ecological restoration measures.

## REFERENCES

- AL-KAISI, M. M.; DOUELLE, A.; KWAW-MENSAH, D. Soil microaggregate and macroaggregate decay over time and soil carbon change as influenced by different tillage systems. **Journal of Soil and Water Conservation**, v. 69, n. 6, p. 574-580, 2014.
- BAYER, C. Armazenamento de carbono em frações lábeis da matéria orgânica de um Latossolo Vermelho sob plantio direto. **Pesquisa Agropecuária Brasileira**, v. 39, n. 7, p. 677-683, 2004.
- BAYER, C.; MARTIN-NETO, L.; MIELNICZUK, J.; CERETTA, C. A. Effect of no-till cropping systems on soil organic matter in a sandy clay loam Acrisol from Southern Brazil monitored by electron spin resonance and nuclear magnetic resonance. **Soil and Tillage Research**, v. 53, p. 95-104, 2000.
- BRANCALION, P. H. S.; VIANI, R. A. G.; RODRIGUES, R. R.; GANDOLFI, S. Avaliação e monitoramento de áreas em processo de restauração. In: MARTINS, S. V. **Restauração ecológica de ecossistemas degradados**. Viçosa: Editora UFV; 2015.
- CAMBARDELLA, C. A.; ELLIOT, E. T. Particulate soil organic-matter changes across a grassland cultivation sequence. **Soil Science Society of American Journal**, v. 56, n. 3, p. 777-783, 1992.
- CORREIA, G. G. S.; MARTINS, S. V. Banco de sementes do solo de floresta restaurada, Reserva Natural Vale, ES. **Floresta e Ambiente**, v. 22, n. 1, p. 79-87, 2015.
- DONAGEMMA, G. K.; CAMPOS, D. V. B.; CALDERANO, S. B.; TEIXEIRA, W. G.; VIANA, J. H. M. **Manual de métodos de análise de solos**. Rio de Janeiro: Embrapa Solos; 2011.
- DORMANN, C. E. Effects of incorporation spatial autocorrelation into the analysis of species distribution data. **Global Ecology and Biogeography**, v. 16, n. 2, p. 129-138, 2007.
- DRIVER, J. D.; HOLBEN, W. E.; RILLIG, M. C. Characterization of glomalin as hyphal wall component of arbuscular mycorrhizal fungi. **Soil Biology and Biochemistry**, v. 37, n. 1, p. 101-106, 2005.
- DURIGAN, G.; FRANCO, G. A. D. C.; SAITO, M.; BAITELLO, J. B. Estrutura e diversidade do componente arbóreo da floresta na Estação Ecológica dos Caetetus, Gália, SP. **Revista brasileira de Botânica**, v. 23, n. 4, p. 371-383, 2000.

- FOKOM, R.; ADAMOU, S.; TEUGWA, M. C.; BEGOUDE BOYOGUENO, A. D.; NANA, W. L.; NGONKEU, M. E. L.; TCHAMENI, N. S.; NWAGA, D.; TSALA NDZOMO, G.; AMVAM ZOLLO, P. H. Glomalin related soil protein, carbon, nitrogen and soil aggregate stability as affected by land use variation in the humid forest zone of south Cameroon. **Soil and Tillage Research**, v. 120, p. 69-75, 2012.
- FUNDAÇÃO SOS MATA ATLÂNTICA. **MG volta a liderar ranking de desmatamento da mata Atlântica: Histórico do desmatamento**. São Paulo, SP: Fundação SOS Mata Atlântica, 2016.
- FUNDAÇÃO SOS MATA ATLÂNTICA/INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE. **Atlas dos Remanescentes Florestais da Mata Atlântica Período 2011-2012**. São Paulo, 2013.
- GERDERMANN, J. W.; NICOLSON, T. H. Spores of mycorrhizal Endogone species extracted from soil by wet sieving and decanting. **Transactions of the British Mycological Society**, v. 46, n. 2, p. 235-244, 1963.
- GREGO, C. R.; VIEIRA, S. R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. **Revista Brasileira de Ciência do Solo**, v. 29, p. 169-177, 2005.
- GUIMARÃES, D. V.; GONZAGA, M. I. S.; MELO NETO, J. O.; REIS, A. F.; LIMA, T. S.; SANTANA, I. L. Qualidade da matéria orgânica do solo e estoques de carbono e nitrogênio em fragmento de Mata Atlântica do município de Neópolis, Sergipe. **Scientia Plena**, v. 8, n. 4, p. 1-5, 2012.
- ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989.
- LANDIM, P. M. B. Sobre Geoestatística e mapas. **Terræ Didática**, v. 2, n. 1, p. 19-33, 2006.
- MENG, Q.; CIESZEWSKI, C.; MADDEN, M. Large area forest inventory using Landsat ETM+: A geostatistical approach. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 64, n. 1, p. 27-36, 2009.
- MIRANDA, C. C.; CANELLAS, L. P.; NASCIMENTO, M. T. Caracterização da matéria orgânica do solo em fragmentos de Mata Atlântica e em plantios abandonados de eucalipto. **Revista Brasileira de Ciência do Solo**, v. 31, n. 5, p. 905-916, 2007.
- MUKAKA, M. M. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. **Malawi Medical Journal**, v. 24, n. 3, p. 69-71, 2012.
- QIN, H.; CHEN, J.; WU, Q.; NIU, L.; LI, Y.; LIANG, C.; SHEN, Y.; XU, Q. Intensive management decreases soil aggregation and changes the abundance and community compositions of arbuscular mycorrhizal fungi in Moso bamboo (*Phyllostachys pubescens*) forests. **Forest Ecology and Management**, v. 400, p. 246-255, 2017.
- R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>
- RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. **geoR: a package for geostatistical analysis**. R NEWS, v. 1, n. 2, p. 15-18, 2001.
- RILLIG, M. C. Arbuscular mycorrhizae, glomalin, and soil aggregation. **Canadian Journal of Soil Science**, v. 84, n. 4, p. 355-363, 2004.
- RODE, R.; FIGUEIREDO FILHO, A.; MACHADO, S. A.; GALVÃO, F. Análise do padrão espacial de espécies e de grupos florísticos estabelecidos em um povoamento de *Araucaria angustifolia* e em uma Floresta Ombrófila Mista no Centro-Sul do Paraná. **Revista Floresta**, v. 40, n. 2, p. 255-268, 2010.
- ROPPA, C.; VALCARCEL, R.; BAYLÃO JUNIOR, H. F. Avaliação da regeneração em ecossistemas perturbados como indicador da restauração em ambientes com marcada estacionalidade, Nova Iguaçu (RJ). **Floresta**, v. 42, n. 2, p. 257-268, 2012.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; CUNHA, T. J. F.; OLIVEIRA, J. B. **Sistema Brasileiro de Classificação de Solos**. Brasília: Embrapa, 2013.

SCHIMIGUEL, R.; SÁ, J. C. M.; BRIEDIS, C.; HARTMAN, D. C.; ZUFFO, J. Estabilidade de agregados do solo devido a sistemas de cultivo. **Synergismus scyentifica**, v. 9, n. 1, p. 1-5, 2014.

SER (Society For Ecological Restoration International E Policy Working Group). **The SER International Primer on Ecological Restoration**. www.ser.org y Tucson: Society for Ecological Restoration International, 2004.

SIGNOR, D.; ZANI, C. F.; PALADINI, A. A.; DEON, M. D.; CERRI, C. E. P. Estoques de carbono e qualidade da matéria orgânica do solo em áreas cultivadas com cana-de-açúcar. **Revista Brasileira de Ciência do Solo**, v. 38, n. 5, p. 1402-1410, 2014.

SILVA, C. S.; PEREIRA, M. G.; DELGADO, R. C.; ASSUNÇÃO, S. A. Spatialization of fractions of organic matter in soil in an agroforestry system in the Atlantic Forest, Brazil. **Cerne**, v. 23, n. 2, p. 249-256, 2017.

SOIL SURVEY STAFF. **Keys to soil taxonomy**. 12th ed. Washington, DC: United States Department of Agriculture, Natural Resources Conservation Service; 2014.

TEDESCO, M. J.; GIANELLO, C.; BISSANI, C. A.; BOHNEN, H.; VOLKWEISS, S. J. **Análise de solos, plantas e outros materiais** (2. ed.). Porte Alegre: Departamento de Solos, UFRGS, 1995.

WEIL, R. R.; ISLAM, K. R.; STINE, M. A.; GRUVER, J. B.; SAMSON-LIEBIG, S. E. Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. **American Journal of Alternative Agriculture**, v. 18, n. 1, p. 3-17, 2003.

WRIGHT, S. F.; UPADHYAYA, A. A survey of soils for aggregate stability and glomalin, a glycoprotein produced by hyphae of arbuscular mycorrhizal fungi. **Plant and Soil**, v. 198, n. 1, 97-107. 1998.

YUN-QIANG, W.; XING-CHANG, Z.; JING-LI, Z.; SHUN-JI, L. Spatial variability of soil organic carbon in a Watershed on the Loess Plateau. **Pedosphere**, v. 19, n. 4, p. 486-495, 2009.

## MODELAGEM ESPACIALIZADA DA EVAPOTRANSPIRAÇÃO REAL EM ÁREA DE REFLORESTAMENTO POR MEIO DO PACOTE AGRIWATER EM AMBIENTE R

Data de aceite: 12/02/2020

### César de Oliveira Ferreira Silva

Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Departamento de Engenharia Rural, Campus de Botucatu-SP.

E-mail: cesaroliveira.f.silva@gmail.com

### Pedro Henrique Jandreice Magnoni

Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Departamento de Engenharia Rural, Campus de Botucatu-SP.

E-mail: pedrohjmagnoni@gmail.com

**RESUMO:** O presente trabalho teve como objetivo modelar espacialmente a evapotranspiração real, assim como outros componentes do balanço de energia no solo, na Estação Experimental de Ciências Florestais de Itatinga, no município de Itatinga – SP, pela aplicação do algoritmo SAFER em ambiente R por meio do pacote *agriwater*. Para tanto foram usados dados remotamente sensoriados associados a dados agrometeorológicos medidos em uma estação automática. Como resultado foram obtidas imagens em arquivos matriciais (rasters) das variáveis albedo superficial, NDVI, temperatura de superfície, saldo de radiação, fração evapotranspirativa e evapotranspiração real modelado espacialmente dentro dos limites da estação.

**PALAVRAS-CHAVE:** Evapotranspiração, pacote *agriwater*, SAFER.

### ACTUAL EVAPOTRANSPIRATION SPATIAL MODELLING IN REFORESTATION AREA USING AGRIWATER PACKAGE IN R

**ABSTRACT:** This work aimed to modelling the actual evapotranspiration and other energy balance components on soil in the Estação Experimental de Ciências Florestais de Itatinga, São Paulo, Brazil. The SAFER algorithm was applied in the R environment using the *agriwater* package. Remote sensing data associated with agrometeorological data measured by an automatic station was used. As result, matrix files (rasterfiles) of surface albedo, NDVI, surface temperature, net radiation, evapotranspiration fraction and actual evapotranspiration was spatially modeled inside the station boundaries. **KEYWORDS:** Evapotranspiration, *agriwater* package, SAFER.

## 1 | INTRODUÇÃO

O processo de transferência da água do solo pela transpiração das plantas é denominado como evapotranspiração. Este fenômeno é um importante componente do ciclo hidrológico já que é responsável pelo

retorno da água precipitada à atmosfera e, portanto, regula a disponibilidade hídrica seja em ambientes naturais ou atividades antrópicas, tais como cultivos agrícolas (BERTONI, 2014). No entanto, métodos tradicionais de medição pontual desta variável são ineficientes em escalas regionais devido a grande variabilidade espaço-temporal, assim como alto custo operacional (SILVA, 2019).

A aplicação de técnicas de sensoriamento remoto permite a obtenção de diferentes variáveis agrometeorológicas de forma espacializada. Dados provenientes do imageamento da superfície terrestre por satélites fornecem informações com diferentes resoluções espaciais e temporais sobre as superfícies vegetadas. Isto auxilia no monitoramento de variáveis biofísicas que impactam nos fluxos de água e energia no solo, como o albedo, biomassa e índice de área foliar (MU et al., 2011, PAIVA et al., 2011).

Modelos para a obtenção da evapotranspiração por meio de sensoriamento remoto envolvem equações empíricas elaboradas pelo cruzamento de dados radiométricos de imagens de satélites com medições em campo da variável que se deseja modelar. O saldo de radiação do solo (componente do balanço de energia) é na em grande parte distribuído como calor latente, calor sensível e fluxo de calor (SILVA, 2019). Diferentes modelos para o balanço de energia são encontrados na literatura como o SEBAL (BASTIAANSSEN et al., 1998a; BASTIAANSSEN et al., 1998b) e METRIC (ALLEN et al., 2005) por meio de equações empíricas, enquanto o SAFER (TEIXEIRA et al., 2010) adiciona também índices de vegetação como o NDVI.

Silva, Teixeira e Manzione (2019) formularam o pacote *agriwater* para aplicação do algoritmo SAFER (TEIXEIRA et al., 2010) em ambiente R. Por meio deste pacote um usuário pode, dentre outros componentes do balanço de energia, obter a evapotranspiração real ( $ET_A$ ) espacializada para imagens dos sensores/satélite OLI/LANDSAT8, MODIS/TERRA, MODIS/AQUA e MSI/SENTINEL2 associadas a dados de uma ou mais estações agrometeorológicas.

Dessa forma, o objetivo deste trabalho foi mostrar a aplicabilidade do algoritmo SAFER (TEIXEIRA et al., 2010) em ambiente R por meio do pacote *agriwater* (SILVA, TEIXEIRA e MANZIONE, 2019), para a determinação de componentes do balanço de energia no solo, assim como a evapotranspiração real, em uma área de reflorestamento com eucaliptos no município de Itatinga – SP.

## 2 | MATERIAL E MÉTODOS

### 2.1 Área de estudo

A modelagem foi realizada dentro dos limites da Estação Experimental de

Ciências Florestais de Itatinga - EECFI, entre os paralelos 23°02'56,96" e 23°03'35,95" latitude sul, e entre os meridianos 48°37'37,83" e 48°37'35,95" longitude oeste, e aproximadamente 22,5 km<sup>2</sup> (Figura 1). Atualmente, sob administração da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), apresenta como utilização da terra majoritariamente o reflorestamento com eucaliptos para diferentes fins de ensino e pesquisa. Também existem fragmentos de vegetação ripária natural junto a corpos hídricos presentes nos limites da estação.

O clima do município de Itatinga é definido como tropical de altitude (Cwa), segundo a classificação de Köppen (1936). Nos meses de verão ocorre maior precipitação acumulada mensal, sendo a média anual de 1372,7 mm. A temperatura do ar média anual é de 19,7°C, com média mínima de 16°C, e média máxima de 22,7°C (CEPAGRI, 2018).

## 2.2 Dados utilizados

As imagens multiespectrais foram obtidas gratuitamente por meio USGS Earth Explorer. Foi adquirida uma imagem Sentinel2B L1C de resolução espacial igual a 10 m, identificada por "*tile number: T22KGV*" (identificação da cena de imageamento) para 09/02/2019, correspondente ao dia juliano (DJ) 40 (ID: L1C\_T22KGV\_A021199\_20190714T132238). Quanto ao sensor MODIS, foi adquirida uma imagem MOD09A1 V6 de resolução espacial igual a 500 m, com a cena identificada pelas coordenadas geográficas lat/long (graus decimais) 25,00336 S e 49,8581 O para o dia 02/02/2019 (DJ 33) (ID: MOD09A1.A2019185.h13v11.006.2019194033408).



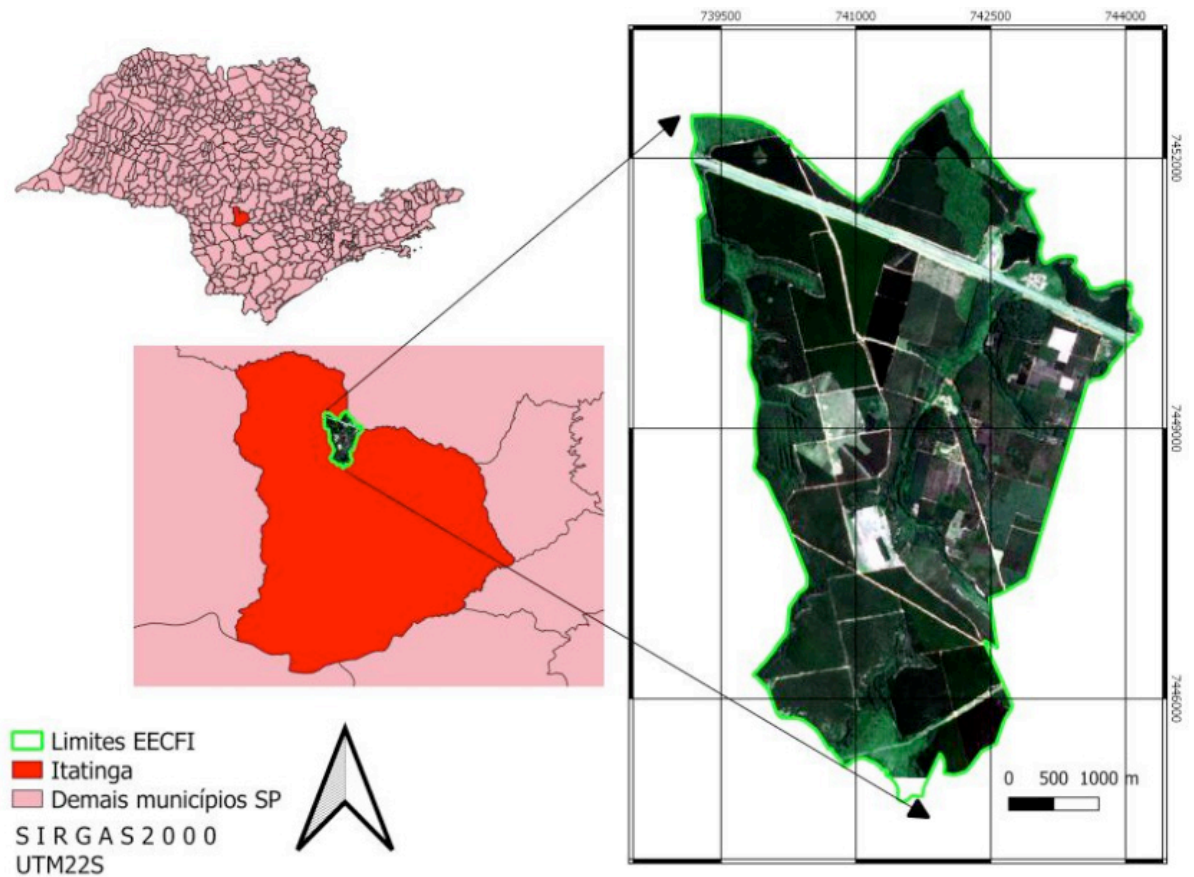


Figura 1. Mapa de localização da Estação Experimental de Ciências Florestais de Itatinga (EECFI).

As bandas espectrais de ambos os sensores foram convertidas de seus formatos de arquivo originais para o formato .tif e seus sistemas de coordenadas foram reprojatados para WGS84.

Uma estação agrometeorológica existente dentro dos limites da EECFI, localizada pela latitude  $23^{\circ} 2' 50,62''$  S e Longitude  $48^{\circ} 38' 18,77''$  O, foi responsável pela aquisição de dados de radiação solar global ( $R_G$ , MJ.m<sup>2</sup>), temperatura média do ar ( $T_A$ , °C) e evapotranspiração de referência ( $ET_o$ , mm). Esses dados foram utilizados conjuntamente com os parâmetros obtidos das imagens multiespectrais na aplicação do SAFER.

### 2.3 Simple Algorithm for Evapotranspiration Retrieving (SAFER)

A razão entre a evapotranspiração real ( $ET_A$ ) e a evapotranspiração de referência ( $ET_o$ ) denominada como fração evapotranspirativa ( $ET_A/ET_o$ ), foi calculada pela Equação 1 (TEIXEIRA, 2010).

$$\frac{ET_A}{ET_o} = \exp \left[ e + f \left( \frac{T_s}{\alpha_0 \cdot NDVI} \right) \right] \quad (1)$$

Onde  $\alpha_0$  corresponde ao albedo da superfície,  $T_s$  a temperatura de superfície, e  $e$  e  $f$  são coeficientes de regressão, sendo 1,8 e -0,008, respectivamente.

A partir da obtenção da fração evapotranspirativa pode-se obter a evapotranspiração real pela Equação 2.

$$ET_A = \frac{ET_A}{ET_0} ET_0 \quad (2)$$

A  $ET_A/ET_0$  em culturas irrigadas, como pivôs centrais, sem estresse ambiental e hídrico, é conhecida como coeficiente de cultura ( $K_c$ ) (ALLEN et al., 1998). Em vegetação natural a fração indica o grau de umidade na zona das raízes das plantas e pode ser usada para caracterizar condições de estresse hídrico (TEIXEIRA et al., 2017, SILVA, TEIXEIRA e MANZIONE, 2019).

A Figura 2 mostra as etapas do cálculo do modelo SAFER onde as caixas com linhas contínuas representam as variáveis espacializadas obtidas por sensoriamento remoto e as tracejadas os dados de entradas da estação agrometeorológica. Onde  $R_N$  é o saldo de radiação,  $T_A$  a temperatura média do ar,  $R_G$  é radiação solar global,  $ET_0$  a evapotranspiração de referência e  $ET_A$  a evapotranspiração real.

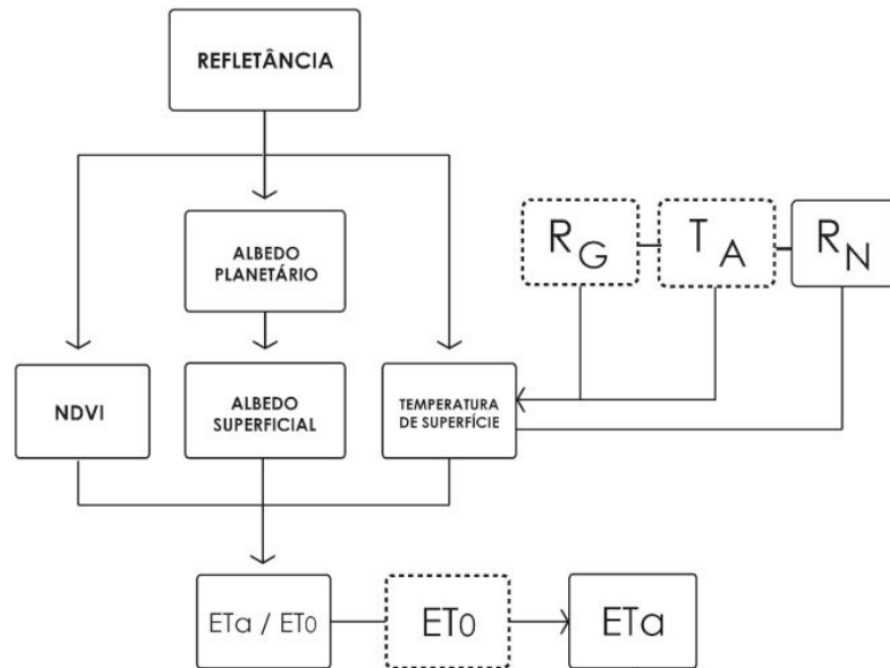


Figura 2. Fluxograma para calcular a evapotranspiração real ( $ET_A$ ) e fração evapotranspirativa ( $ET_A/ET_0$ ) pelo modelo SAFER sem o uso da banda termal. (SILVA, TEIXEIRA e MANZIONE, 2019).

A seguir as equações de modelagem espacial dos dados de sensoriamento remoto para as imagens dos sensores OLI/LANDSAT8, MODIS/TERRA&AQUA e MSI/SENTINEL2 serão detalhadas segundo a metodologia de Silva, Teixeira e Manzione, 2019.

## 2.4 Albedo Superficial

Para determinar o albedo superficial ( $\alpha_0$ ), inicialmente o albedo planetário para

todo o espectro solar ( $\alpha_p$ ) foi calculado como a soma total das diferentes refletâncias de banda ( $r_{band}$ ) de acordo com os pesos de cada banda ( $w_{band}$ ) de acordo com a Equação 3.

$$\alpha_p = \sum w_{band} r_{band} \quad (3)$$

Na utilização de imagens do sensor MODIS, no qual o albedo planetário é calculado mediante a Equação 4.

$$\alpha_p = a + \sum w_{band} r_{band} \quad (4)$$

Os pesos das bandas espectrais dos sensores são encontrados pela razão entre a quantidade de radiação de onda curta de cada banda e a soma da radiação de onda curta recebida para todas as bandas no topo da atmosfera (TOA) (Tabela 1).

Sensor	Banda	Coefficiente	Sensor	Banda	Coefficiente
OLI	1	0,10	MSI	2	0,32
	2	0,31		3	0,26
	3	0,30		4	0,25
	4	0,13		8	0,17
	5	0,08	MODIS	a	0,08
	6	0,05		1	0,41
	7	0,04		2	0,14

Tabela 1. Pesos para correção atmosférica no topo da atmosfera para as diferentes bandas espectrais dos sensores MSI/SENTINEL2, OLI/LANDSAT8, MODIS/TERRA&AQUA (Adaptado de SILVA, TEIXEIRA e MANZIONE, 2019).

Finalmente, os valores diários do albedo superficial são obtidos pela Equação 5.

$$\alpha_0 = b\alpha_p + c \quad (5)$$

onde  $b$  e  $c$  são coeficientes de regressão, que para um período de 24 horas foram considerados, respectivamente, como 1,70 e 0,13 (TEIXEIRA et al., 2008; TEIXEIRA et al., 2014a; TEIXEIRA et al., 2014b).

## 2.5 NDVI

O Normalized Difference Vegetation Index (NDVI) (ROUSE et al., 1974) foi calculado pela Equação 6.

$$NDVI = \frac{r_{NIR} - r_{RED}}{r_{NIR} + r_{RED}} \quad (6)$$

onde  $\tau_{NIR}$  e  $\tau_{RED}$  são as reflectâncias na faixa de comprimento de onda do infravermelho próximo (NIR) e do vermelho (RED).

## 2.6 Temperatura superficial

Para modelar a temperatura superficial sem uso das bandas termais dos sensores imageadores foi utilizado o método do resíduo da equação de Stefan-Boltzmann (TEIXEIRA et al., 2016), seguindo a Equação 7.

$$T_S = \sqrt[4]{\frac{\varepsilon_A \sigma T_A^4 + \alpha_L \tau_{sw}}{\varepsilon_S \sigma}} \quad (7)$$

onde  $T_A$  é a temperatura média do ar da estação agrometeorológica dentro da área de estudo com índice de área foliar (IAF) de 2,88 e albedo estimado de 0,23,  $\tau_{sw}$  é a transmissividade atmosférica de ondas curtas, definidas com 44% da  $R_G$  no topo da atmosfera (TEIXEIRA et al., 2008, 2010),  $\varepsilon_A$  e  $\varepsilon_S$  são, respectivamente, as emissividades atmosférica e superficial e  $\sigma$  é a constante de Stefan-Boltzmann ( $5,67 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ ).  $\varepsilon_A$  e  $\varepsilon_S$  são calculadas pelas equações 8 e 9, respectivamente (TEIXEIRA et al. 2008 e 2014a).

$$\varepsilon_A = a_A (-\ln \tau_{sw})^{b_A} \quad (8)$$

$$\varepsilon_S = a_S \ln NDVI + b_S \quad (9)$$

onde  $a_A$ ,  $b_A$ ,  $a_S$  e  $b_S$  são coeficientes de regressão, respectivamente, 0,94, 0,10, 0,06 e 1,00 (TEIXEIRA et al., 2010). Teixeira et al., (2009) calibraram  $\varepsilon_A$  e  $\varepsilon_S$  com  $R^2$  de 0,75 e 0,90 respectivamente. Finalmente, o coeficiente  $\alpha_L$  é dado pela Equação 10.

$$\alpha_L = cT_A - d \quad (10)$$

onde  $c$  e  $d$  são coeficientes de regressão iguais a 6,99 e 39,93 (TEIXEIRA et al. 2014a).

## 2.7 Balanço de energia no solo

O saldo de radiação ( $R_N$ ,  $\text{MJ m}^{-2} \text{ dia}^{-1}$ ) pode ser obtido segundo a Equação 11 derivada por Teixeira et al. (2008).

$$R_N = (1 - \alpha_0)R_G - \alpha_L \tau_{sw} \quad (11)$$

onde  $\tau_{sw}$  é a transmissividade atmosférica de ondas curtas definida como a razão de  $R_G$  para a radiação solar incidente no topo da atmosfera.

O balanço de energia considerado pelo SAFER é descrito na Equação 12.

$$R_N = \lambda E + H + G \quad (12)$$

onde  $R_N$  é o saldo de radiação,  $\lambda E$  é o fluxo de calor latente,  $H$  é o fluxo de calor sensível e  $G$  é o fluxo de calor no solo.

Para obter-se o fluxo de calor latente ( $\lambda E$ ), aplica-se a Equação 13.

$$\lambda E = 2,45 ET_A \quad (13)$$

O fluxo de calor no solo ( $G$ ) em escala diária é obtido pela Equação 14.

$$G = R_N g \exp(h \alpha_0) \quad (14)$$

onde  $g$  e  $h$  são coeficientes de regressão que foram considerados, respectivamente, 3,98 e -31,98 (TEIXEIRA et al., 2008).

Finalmente,  $H$  é estimado como resíduo do balanço de energia na Equação 15.

$$H = R_N - \lambda E - G \quad (15)$$

## 2.8 Agriwater

O download do *agriwater* pode ser realizado através do repositório CRAN no link <https://cran.r-project.org/web/packages/agriwater/index.html> ou instalado diretamente no R com a função “install.packages()”. Para utilizar o *agriwater*, os pacotes *sp*, *raster* e *rgdal* também devem ser instalados. A rotina utilizada no R para as análises pode ser acessada e baixada no seguinte link: [doi.org/10.13140/RG.2.2.34818.84163](https://doi.org/10.13140/RG.2.2.34818.84163)

Após os pacotes devidamente instalados, os dados de entrada são uma máscara *shapefile* da área de estudo mais as informações como visto na figura 2. Todos os arquivos devem estar em um mesmo diretório que deverá ser indicado no R. Dentro do pacote, existem funções específicas para cada satélite. Neste estudo foram usadas as funções “evapo\_s2” e “evapo\_modis” (Tabela 2). Existem diversas outras funções dentro do pacote que podem ser vistas em Silva, Teixeira e Manzione, 2019. Mais exemplos práticos de aplicações do pacote *agriwater* podem ser reproduzidas a partir da leitura de Leivas et al. (2019).

Funções	Breve descrição	Dados de entrada	Rasters gerados
evapo_s2	Modelagem da evapotranspiração real ( $ET_a$ )	Temperatura média do ar, radiação solar e evapotranspiração de referência de uma única estação agrometeorológica e imagens digitais	Alb_24.tif, NDVI.tif, LST.tif, Rn_MJ, kc.tif and evapo.tif
evapo_modis			

Tabela 8. Descrição, dados de entrada e saídas do comando utilizado no pacote *agriwater* para balanço de energia e recuperação de evapotranspiração real. “s2” no nome do comando indica que ele usa imagens de sensores/satélites MSI/SENTINEL2, “modis” o uso de MODIS/TERRA&AQUA.

### 3 | RESULTADOS E DISCUSSÃO

A aplicação do pacote *agriwater* aos dados disponíveis resultou na geração de arquivos *raster* de extensão *.tif* referentes ao albedo superficial em escala diária (Figura 3), NDVI (Figura 4), temperatura de superfície em Kelvin (K) (Figura 5), saldo de radiação em MJ/m<sup>2</sup>.dia (Figura 6), fração evapotranspirativa (Figura 7) e evapotranspiração real em mm (Figura 8). Todos os resultados foram plotados a partir dos pacotes *rasterVis* e *gridExtra* no R.

Em todos os produtos resultantes fica evidente o efeito da resolução espacial. Por apresentar baixa resolução espacial (500 m), as variáveis modeladas na imagem do sensor MODIS apresentam maior mistura espectral nos *rasters* gerados, fator que diminui a sensibilidade dos valores estimados e atrapalha a interpretação dos mapas gerados. Isso demonstra a importância da escolha correta da resolução espacial da imagem a ser modelada levando em conta o tamanho da área de estudo. Na Figura 8, ao norte da EECFI, é possível observar os valores iguais a zero de evapotranspiração real (mm) no trecho da Rodovia Castelo Branco que corta a estação na modelagem baseada no sensor MSI (resolução espacial de 10 m). Na modelagem para o MODIS os valores de  $ET_A$  na rodovia são diluídos pela resolução espacial (Figura 8)

Pela falta de dados medidos em campo da  $ET_A$  não foi possível validar os dados de evapotranspiração modelados. No entanto, outros trabalhos já demonstraram aplicações bem-sucedidas do SAFER em demais localidades como o Pantanal (TEIXEIRA et al., 2015), Rio Grande do Sul (LEIVAS et al., 2015) e outras localidades de São Paulo (HERNANDEZ et al., 2014).

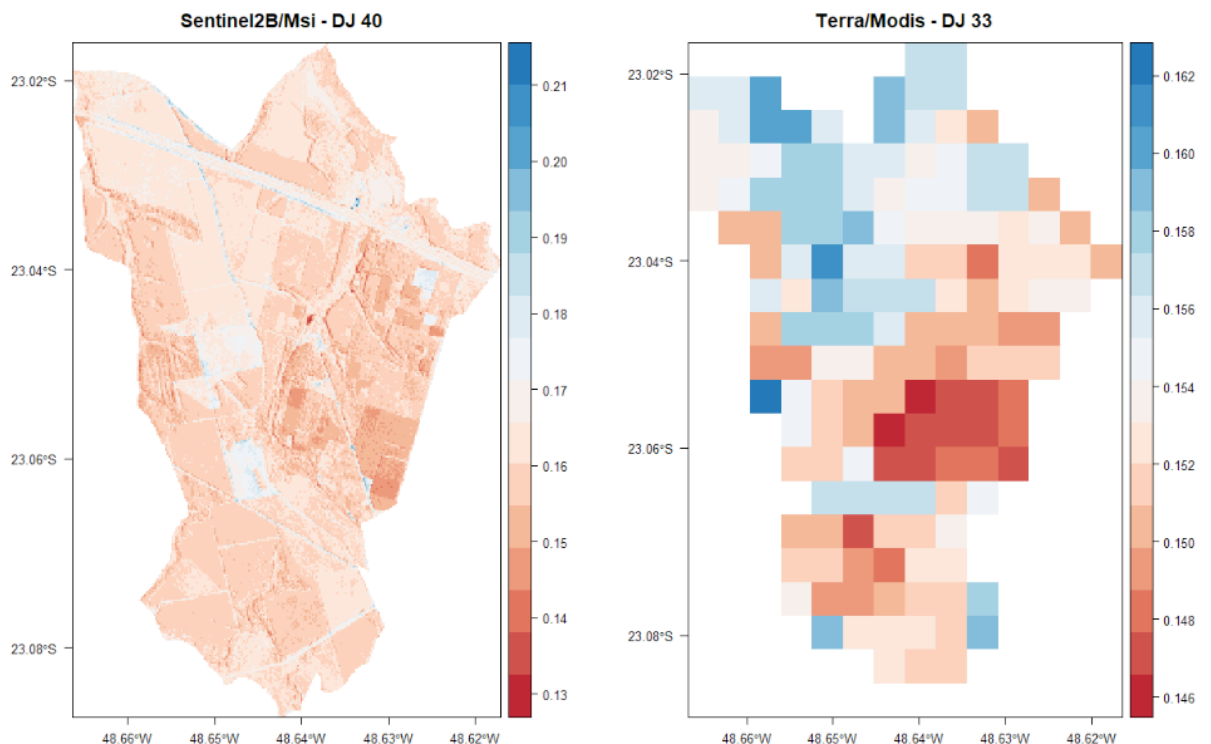


Figura 3. Albedo superficial modelado para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.



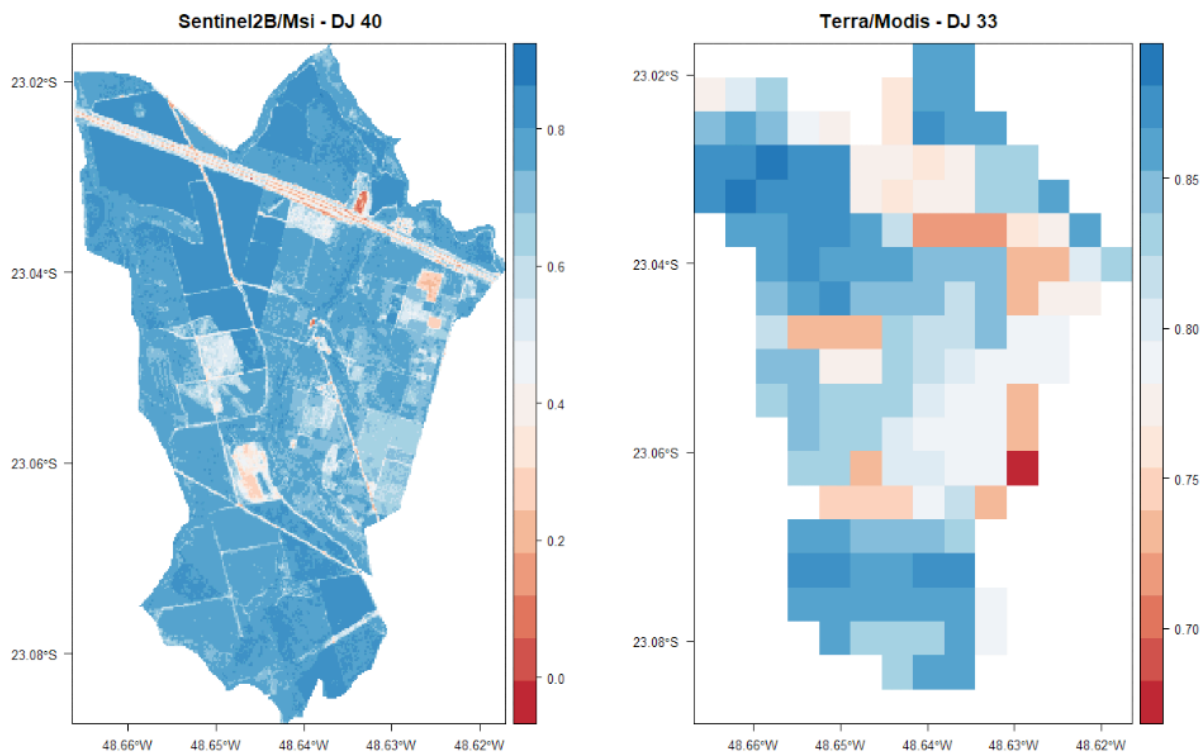


Figura 4. NDVI calculado para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.

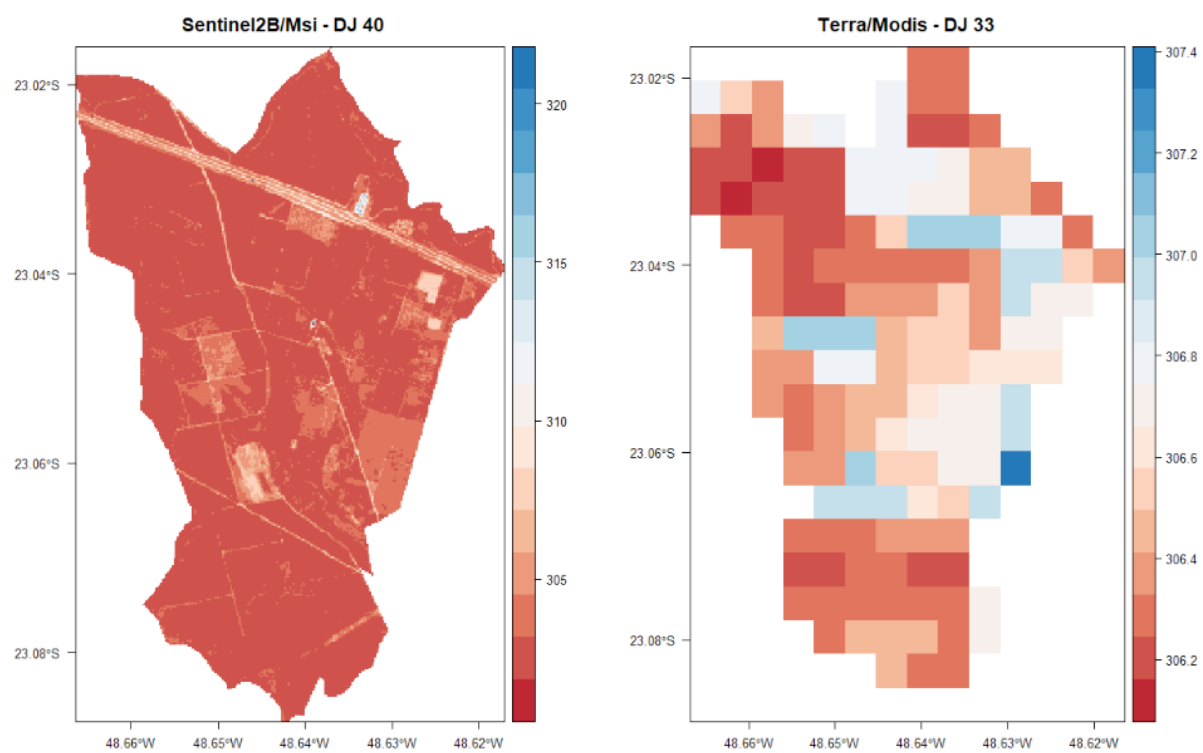


Figura 5. Temperatura de superfície (K) modelada para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.

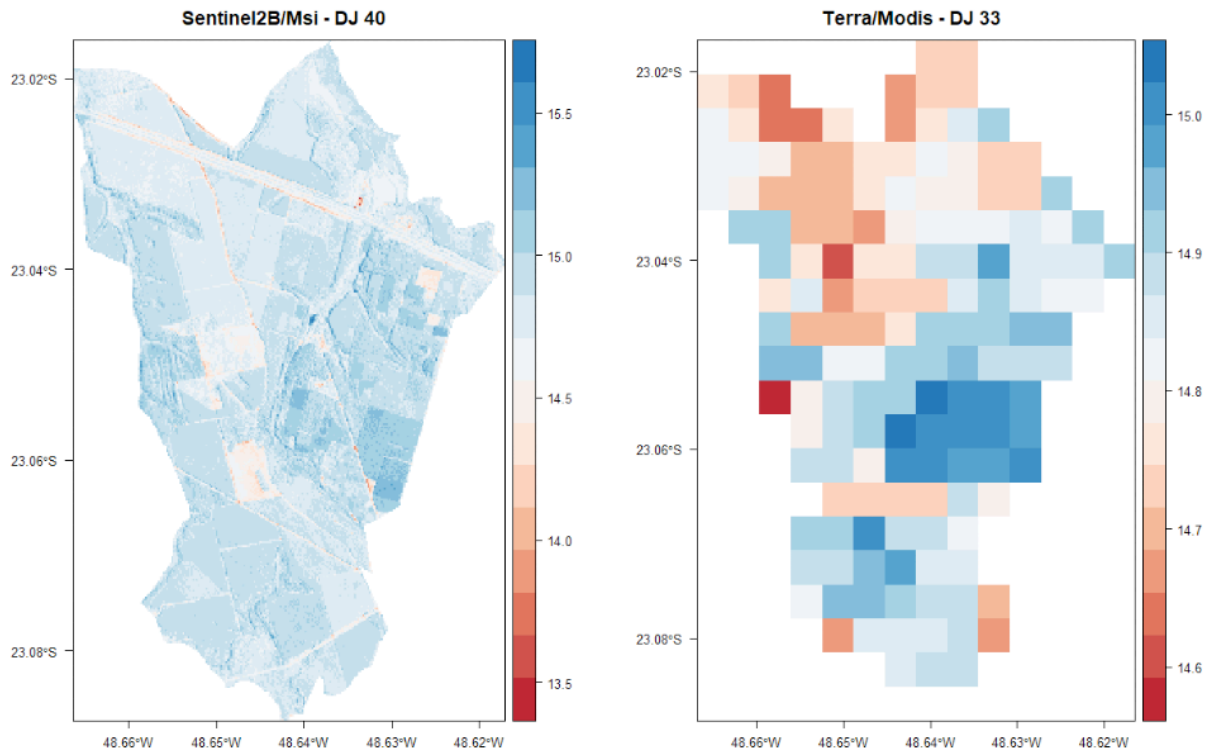


Figura 6. Saldo de radiação (MJ/m<sup>2</sup>.dia) modelado para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.

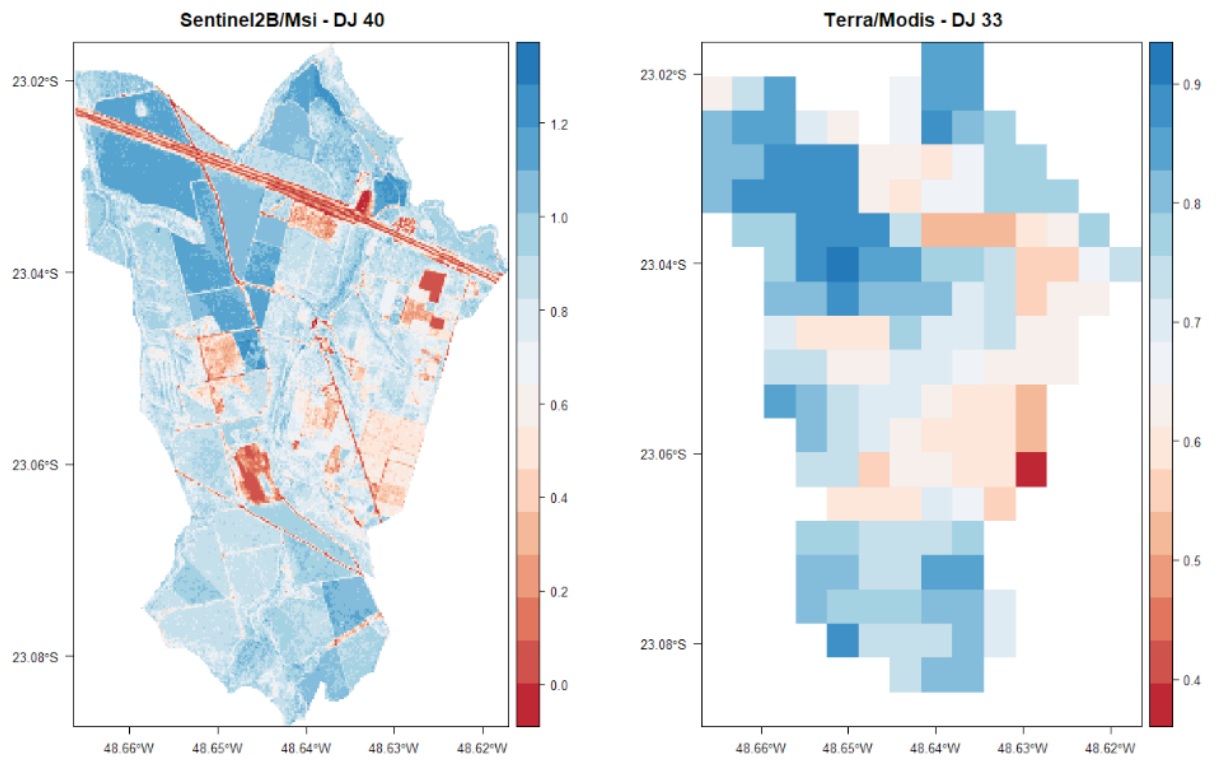


Figura 7. Fração evapotranspirativa modelada para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.

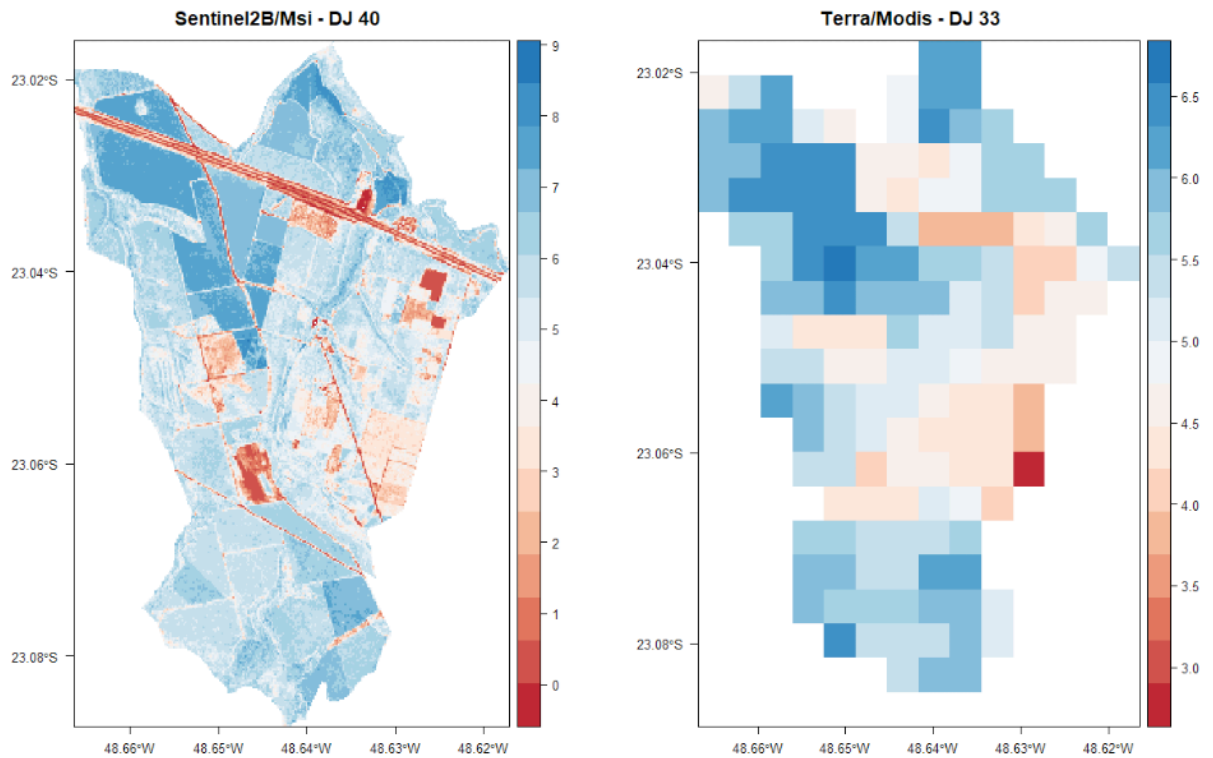


Figura 8. Evapotranspiração real (mm) modelada para MSI/SENTINEL2 no DJ40 e TERRA/MODIS no DJ 33 na EECFI.

## 4 | CONCLUSÕES

Por se tratar de um ambiente de programação de código aberto, o R é um software que está ganhando rapidamente o uso em diferentes áreas do conhecimento. É uma linguagem que oferece recursos de análise interativa e é ideal para o rápido desenvolvimento de aplicativos estatísticos e de análise de dados. O pacote *agriwater* é um exemplo, visto sua robustez para a modelagem de dados espaciais na estimativa de parâmetros agrometeorológicos, que auxiliam no entendimento de processos como o ciclo hidrológico e o balanço hídrico. No entanto, é necessário realizar validações em campo dos resultados que avaliem as representatividades em diferentes condições climáticas regionais.

## AGRADECIMENTOS

Os autores agradecem a direção da Estação Experimental de Ciências Florestais de Itatinga (EECFI) pelos dados agrometeorológicos cedidos para o desenvolvimento deste trabalho.

## REFERÊNCIAS

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. **Crop evapotranspiration: guidelines for computing crop water requirements**. FAO, Irrigation and Drainage Paper, 56. Roma, 1998. 300p.

ALLEN, R. G.; WALTER, I. A.; ELLIOTT, R. L.; HOWELL, T. A.; ITENFISU, D.; JENSEN M. E.; SNYDER, R. L. **ASCE Standardized Reference Evapotranspiration Equation**. Task Committee on Standardization of Reference Evapotranspiration – Environmental and Water Resources Institute of the American Society of Civil Engineers, 2005.

BASTIAANSEN, W. G. M.; MENENTI, M.; FEDDES, R. A.; HOLTSLAG, A. M. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. **Journal of Hydrology**, v. 212–213, n. 1-4, p. 198-212. 1998a.

BASTIAANSEN, W. G. M.; PELGRUM, H.; WANG, J.; MORENO, J. F.; ROERINK, G. J.; WAL, T. V. A remote sensing surface energy balance algorithm for land (SEBAL). 2. Validation. **Journal of Hydrology**, v. 212–213, n. 1-4, p. 213-229. 1998b.

BERTONI, J.; LOMBARDI NETO, F. **Conservação do solo**. São Paulo: Ícone, 6ª Ed. 2014.

CEPAGRI. Centro de Pesquisas Meteorológicas e Climáticas Aplicadas a Agricultura. **Clima dos Municípios Paulistas**. São Paulo: Universidade Estadual de Campinas, UNICAMP, 2018.

HERNANDEZ, F. B. T.; NEALE, C. M. U.; TEIXEIRA, A. H. C.; TAGHVAEIAN, S. Determining large scale actual evapotranspiration using agrometeorological and remote sensing data in the northwest of São Paulo state, Brazil. **Acta Horticulturae**, v.1038, p.263-270, 2014.

LEIVAS, J.F.; TEIXEIRA, A.H. C.; ANDRADE, R.G.; VICTORIA, D. de C.; SILVA, G.B.; BOLFE, E.L. Application of agrometeorological spectral model in rice area in southern Brazil. **Proceedings of SPIE**, v. 9637, p. 96372B-1-96372B-8, 2015.

LEIVAS, J. F.; SILVA, C. O. F.; MANZIONE, R. L.; TEIXEIRA, A.H. de C. **Tutorial do pacote Agriwater para modelagem espacial da evapotranspiração no software R**. Campinas: Embrapa Territorial, 2019. 49 p.

MU, Q.; ZHAO, M.; RUNNING, S. W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. **Remote Sensing of Environment**, v. 115, n. 8, p. 1781-1800, 2011.

PAIVA, C. M.; TSUKAHARA, R. Y.; FRANÇA, G. B.; NICACIO, R. M. Estimativa da evapotranspiração via sensoriamento remoto para fins de manejo de irrigação. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 15. (SBSR), 2011.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with ERTS. In: Earth Resources Technology Satellite Symposium, 3., 1973, Washington. **Proceedings**. Washington: NASA, 1974. v. 1, p. 309-317.

SILVA, C. O. F. **MODELAGEM ESPACIAL DA RECARGA DAS ÁGUAS SUBTERRÂNEAS SOB DIFERENTES USOS E COBERTURAS DA TERRA**. Orientador: Rodrigo Lilla Manzione. 2019. Dissertação (Mestrado) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2019. Disponível em: <https://repositorio.unesp.br/handle/11449/190710>. Acesso em: 12 dez. 2019.

SILVA, C. O. F.; TEIXEIRA, A. H. C.; MANZIONE, R. L. Agriwater: An R package for spatial modelling of energy balance and actual evapotranspiration using satellite images and agrometeorological data. **Environmental Modelling & Software**, 2019.

TEIXEIRA, A. H. C.; BASTIAANSEN, W. G. M.; AHMAD, M.; BOS, M.G. Reviewing SEBAL input parameters for assessing evapotranspiration and water productivity for the Low-Middle São Francisco River basin, Brazil Part B: Application to the regional scale. **Agricultural and Forest Meteorology**, v. 149, p. 477-490, 2009.

TEIXEIRA, A. H. C.; HERNANDEZ, F.B.T.; LOPES, H.L.; SCHERER-WARREN, M.; BASSOI, L.H. A Comparative Study of Techniques for Modeling the Spatiotemporal Distribution of Heat and Moisture

Fluxes in Different Agroecosystems in Brazil. In: George G. Petropoulos. (Org.). **Remote Sensing of Energy Fluxes and Soil Moisture Content**. 1ed. Boca Raton, Florida: CRC Group, Taylor and Francis, 2014a, p. 169-191.

TEIXEIRA, A. H. C.; LEIVAS, J. F.; RONQUIM, C. C.; SILVA, G. B.; VICTORIA, D. de C. Energy balances in sugar cane, coffee and natural vegetation in the northeastern side of the São Paulo state, Brazil. **Proceedings of SPIE**, v.9998, p.99980M-1 - 99980M-16, 2016.

TEIXEIRA, A. H. C.; REIS, B. R. S.; LEIVAS, J. F.; SILVA, G. B. S.; STRUIVING, T. B. Componentes da produtividade da água modelados por sensoriamento remoto em limoeiros irrigados de Minas Gerais. **Agrometeoros**, 2017, 25.1.

TEIXEIRA, A. H. C.; PADOVANI, C. R.; ANDRADE, R. G.; LEIVAS, J. F.; VICTORIA, D. DE C.; GALDINO, S. Use of MODIS images to quantify the radiation and energy balances in the Brazilian Pantanal. **Remote Sensing**, v. 7, p. 14597-14619, 2015.

TEIXEIRA, A.H.C., HERNANDEZ, F.B.T.; ANDRADE, R.G.; LEIVAS, J.F.; BOLFE, E.L. Energy balance with Landsat images in irrigated central pivots with corn crop in the São Paulo State, Brazil. **Proceedings of SPIE**, v. 9239, p. 923900-1-923900-10, 2014b.

TEIXEIRA, A.H.C. Determining regional actual evapotranspiration of irrigated and natural vegetation in the São Francisco river basin (Brazil) using remote sensing an Penman-Monteith equation. **Remote Sensing**, v. 2, p. 1287-1319, 2010.

TEIXEIRA, A.H.C.; BASTIAANSSEN, W.G.M.; AHMAD, M.D., MOURA, M.S.B.; BOS, M.G. Analysis of energy fluxes and vegetation-atmosphere parameters in irrigated and natural ecosystems of semi-arid Brazil. **Journal of Hydrology**, v. 362, p. 110-127, 2008.

## IMPACTO DO FOGO NO BANCO DE SEMENTES DE FLORESTA ESTACIONAL SEMIDECIDUAL ALTOMONTANA NO QUADRILÁTERO FERRÍFERO, MG

*Data de aceite: 12/02/2020*

### Júnia Maria Lousada

Universidade Federal de Viçosa, Departamento de Botânica  
Viçosa, MG

### Pedro Manuel Villa

Universidade Federal de Viçosa, Departamento de Botânica  
Viçosa, MG

### Gustavo Heringer

Universidade Federal de Viçosa, Departamento de Botânica  
Viçosa, MG

Universidade Federal de Lavras, Programa de Pós-graduação em Ecologia Aplicada  
Lavras, MG

### Sebastião Venâncio Martins

Universidade Federal de Viçosa, Departamento de Engenharia Florestal  
Viçosa, MG

**RESUMO:** Incêndios de origem antrópica constituem uma das principais ameaças à integridade ambiental no Quadrilátero Ferrífero (QF), região de encaves entre floresta e campo. Avaliamos o potencial de regeneração natural de uma floresta estacional após distúrbio causado pelo fogo através do monitoramento da diversidade e estrutura da vegetação representada no banco de sementes do solo

(BSS). O BSS de uma área adjacente também foi avaliado para servir como referência. Em cada área amostramos 0,5 ha para a coleta do BSS. As hipóteses estabelecidas foram: 1) O fogo reduz a diversidade de espécies do BSS e 2) O fogo provoca a redução de espécies arbóreas e a dominância de gramíneas competidoras no BSS. Nossos resultados apontaram o comprometimento da resiliência de ambas as áreas, sendo a área afetada pelo fogo, a mais crítica, com elevada abundância de herbáceas competidoras (apenas 3 espécies corresponderam a 70,20 % do total de indivíduos amostrados) e a baixa abundância de espécies e indivíduos de arbóreas (3 espécies e 2,66 % do total de indivíduos, respectivamente). Esse resultado foi corroborado pela análise de NMDS que nos mostrou menor variabilidade na composição florística entre as parcelas da área queimada, tanto para composição de espécies quanto para estrutura de indivíduos. Entre as espécies arbóreas da área queimada, apenas uma era zoocórica. A elevada presença de gramíneas competidoras contribui para potencialização de novos distúrbios pelo fogo e dificulta o processo de sucessão secundária florestal, facilitando o processo de savanização de florestas.

**PALAVRAS-CHAVE:** Ecologia do fogo, regeneração natural, resiliência, savanização de florestas.



## IMPACT OF FIRE ON SEED BANCK OF SEASONAL SEMIDECIDUOUS ALTOMONTANA FOREST IN QUADRILÁTERO FERRÍFERO, MG

**ABSTRACT:** Fires of anthropic origin constitute one of the main threats to environmental integrity in the Iron Quadrangle (IQ), a region of forest-field enclaves. We evaluated the natural regeneration potential of a seasonal forest after fire disturbance by monitoring the diversity and structure of the vegetation represented in the soil seed bank (SSB). The SSB from an adjacent area has also been evaluated for reference. In each area we sampled 0.5 ha for SSB collection. The hypotheses established were: 1) Fire reduces species diversity of SSB and 2) Fire causes reduction of tree species and dominance of competing grasses in SSB. Our results pointed to the resilience compromise of both areas, being the most critical area affected by fire, with high abundance of competing herbaceous (only 3 species corresponded to 70.20% of the total sampled individuals) and the low abundance of tree species and individuals (3 species and 2.66% of the total individuals, respectively). This result was corroborated by the NMDS analysis that showed us less variability in floristic composition between the plots of the burned area, both for species composition and for individual structure. Among the tree species in the burned area, only one was zoochoric. The high presence of competing grasses contributes to the potentialization of new disturbances by fire and hinders the process of secondary forest succession, facilitating the process of forest savannization.

**KEYWORDS:** Fire ecology, natural regeneration, resilience, forest savannization.

### 1 | INTRODUÇÃO

Incêndios de origem antrópica constituem uma das principais ameaças à integridade ambiental no Quadrilátero Ferrífero (QF), região de destaque econômico e ambiental no Brasil. Situado na porção sul da Cadeia do Espinhaço, em Minas Gerais, o QF constitui divisor entre os domínios da Mata Atlântica e do Cerrado, (MINAS GERAIS, 2009), apresentando vários encraves vegetacionais entre floresta e campo.

O fogo é uma das principais forças evolutivas que influencia a distribuição, diversidade e estrutura das comunidades vegetais em diferentes tipos de ecossistemas (BOND & KEELEY, 2005; HARDESTY *et al.*, 2005). Por este motivo, compreender como atuam os controladores ambientais de florestas tropicais e das savanas é questão fundamental para lidar com as mudanças futuras na distribuição desses ecossistemas (MURPHY & BOWMAN, 2012). As savanas são ecossistemas que coevoluíram com o fogo e apresentam adaptações que conferem diferentes mecanismos de resistência. Pelo contrario, as florestas tropicais são ambientes sensíveis ao fogo, apresentando pouca ou nenhuma adaptação (HARDESTY *et al.*, 2005).

O solo é uma fonte importante de sementes viáveis que se relaciona

diretamente ao estabelecimento de populações de plantas, de grupos ecológicos, à manutenção da diversidade de espécies e à regeneração natural após distúrbios naturais e antrópicos (HARPER, 1977), sendo um dos mecanismos de regeneração pós-fogo. A composição e estrutura de espécies em um banco de sementes do solo, permitem avaliar, pelo método de germinação, a densidade e riqueza de espécies arbustivo-arbóreas verificando se há comprometimento da resiliência em um ambiente florestal (MARTINS *et al.*, 2012). A síndrome de dispersão mais frequente em florestas tropicais é a zoocórica, com ocorrência mínima de 50% (HOWE & SMALLWOOD, 1982). Em áreas de fogo frequente algumas plantas lenhosas reduzem severamente sua produção de sementes, alterando o banco de sementes populacional. Nestas condições, ocorre aumento de gramíneas e de seu banco de sementes, o que representa um filtro biótico para o estabelecimento de espécies lenhosas (HENRIQUES, 2005).

O fogo frequente diminui a altura da vegetação, altera a composição de espécies, sobretudo de lenhosas sensíveis ao fogo, favorece as arbustivas em detrimento das arbóreas e provoca elevada mortalidade de plântulas e de rebrotas vegetativas (BOND & KEELEY, 2005; HENRIQUES, 2005). Algumas espécies arbóreas do Cerrado apresentam adaptações de casca para resistir ao fogo, mas não apresentam adaptações para seus frutos, comprometendo a reprodução sexual (MIRANDA *et al.*, 2002). Desta forma, o fogo pode promover alterações na composição e estrutura das espécies e influenciar a dinâmica de distribuição de florestas e savanas.

Em nossa pesquisa avaliamos duas áreas de Floresta Estacional Semidecidual (FES) inseridas em contexto de distúrbios antigos e atuais, causados por fogo, em um mosaico entre floresta, campo gramíneo, campo rupestre e cerrado. Nossas hipóteses lançadas foram: 1) O fogo reduz a diversidade de espécies do BSS e 2) O fogo provoca redução de arbóreas e a dominância de gramíneas competidoras.

## 2 | MATERIAL E MÉTODOS

### 2.1 Área de estudo

Realizamos a pesquisa no Monumento Natural Estadual Itatiaia (MNEI), localizada nos municípios de Ouro Preto e Ouro Branco, região do Quadrilátero Ferrífero, Minas Gerais (Figura 1). O MNEI possui área total de 3.216 ha, o clima local é o tropical de altitude (Cwb mesotérmico, classificação de Koeppen).

A temperatura anual média é de 20,7°C e a precipitação média anual é de 1.188,2 mm (MINAS GERAIS, 2009). O relevo é declivoso, com altitudes variando entre 800 e 1400 m. As principais classes de solos encontradas são os Cambissolos, Neossolos litólicos e Latossolos vermelho amarelo. Apresenta uma rica hidrografia

pertencente a duas importantes bacias, a do Rio São Francisco e do Rio Doce, sendo importante fonte de abastecimento da região. A vegetação é composta por um mosaico vegetacional de Floresta Estacional Semidecidual, Campo Rupestre e Cerrado (MINAS GERAIS, 2009).

## 2.2 Coleta dos dados

Amostramos duas áreas de Floresta Estacional Semidecidual Altomontana (FES) localizadas em área de encrave entre FES e Campo gramíneo. Uma foi a área queimada em outubro de 2014, onde o incêndio comprometeu grande parte do dossel do fragmento amostrado, com morte de vários indivíduos arbóreos e total consumo da biomassa arbustiva e herbácea. E a outra área não apresenta sinais de queima recente e será utilizada como ecossistema de referência (Figura 1). Essa área encontram-se em estágio inicial de sucessão secundária florestal (conforme Resolução nº 392 do Conselho Nacional do Meio Ambiente – CONAMA, de 25 de junho de 2007), com presença de dossel irregular, sub-bosque, presença de fina camada de serapilheira, gramíneas e lianas em alguns pontos. Ambas as áreas fazem parte de um grande fragmento florestal (aproximadamente 1.500 ha) e apresentam características semelhantes de estrutura arbórea, declividade e localização, situadas na borda do fragmento, distando 500 m uma da outra. Em outubro de 2014, fizemos o campo de coleta do banco de sementes do solo (BSS) da área afetada pelo fogo e em abril de 2015 coletamos o BSS da área referência.

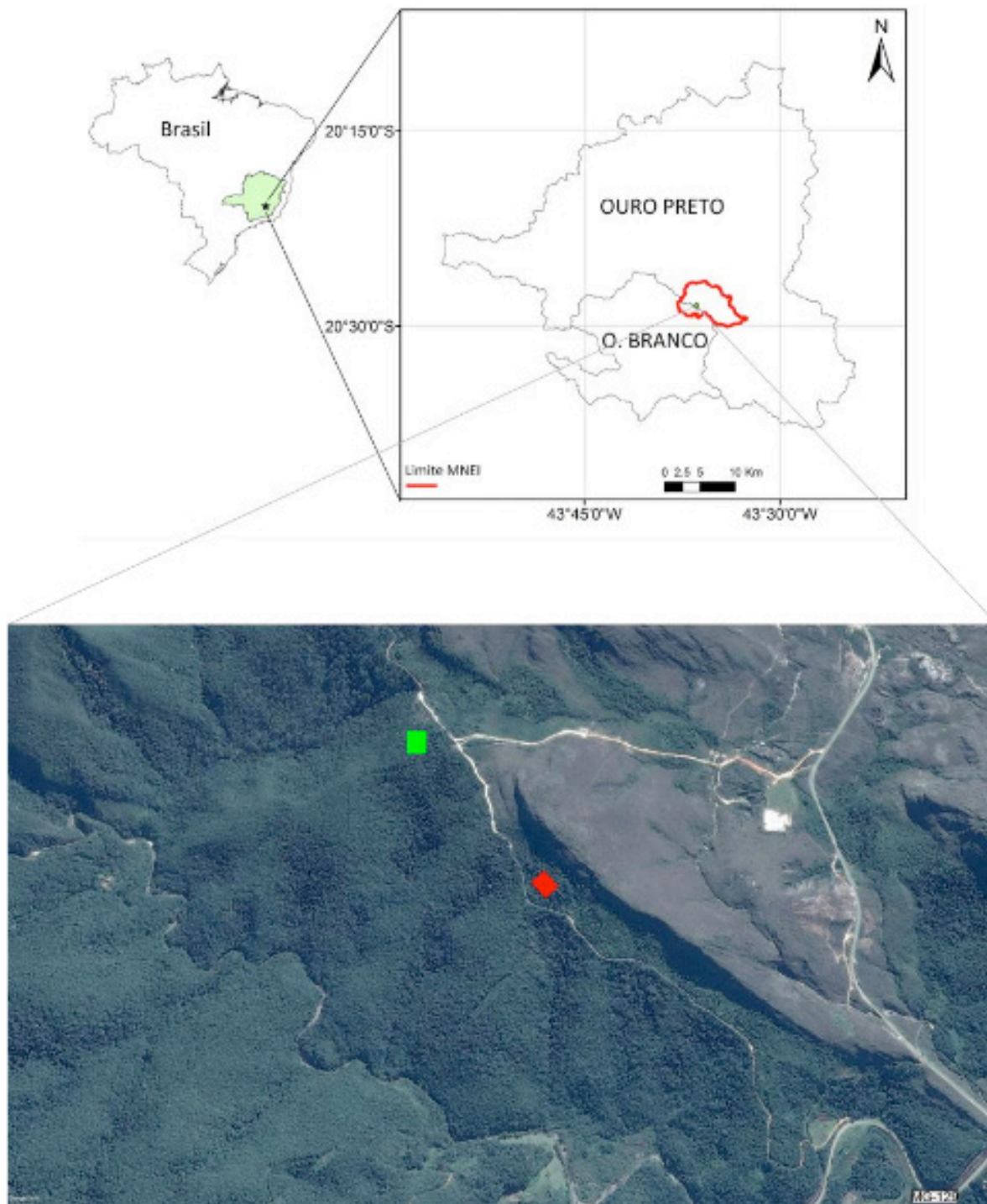


Figura 1. Limite do Monumento Natural Estadual do Itatiaia (MNEI) (em vermelho no mapa) situado nos municípios de Ouro Preto e Ouro Branco, Minas Gerais, Brasil. Imagem satélite com a localização das áreas amostradas, área referência (verde) e área queimada (vermelho). Destaque para o encrave vegetacional entre Floresta Estacional Semidecidual AltoMontana e Campo Rupestre. Mapa: Júnia M. Lousada. Imagem: Google Earth.

Em ambas as áreas amostramos 0,5 ha representado por 25 parcelas de 10 x 10 m, de forma sistemática. No centro de cada parcela lançamos uma moldura de madeira de 40 cm x 30 cm x 5 cm (0,006 cm<sup>3</sup>) para a coleta do BSS.

Levamos as amostras de solo para o viveiro da Universidade Federal de Viçosa, na casa de sombra com tela de 50% de sombreamento e irrigação diária. A identificação e quantificação das plântulas foi realizada quinzenalmente durante um

ano.

Classificamos as espécies dentro de três grupos ecológicos; forma de vida, dispersão de sementes e indicadores do estágio de regeneração. Para formas de vida consultamos o Herbário Virtual Re flora (arbóreas, arbustivas, ervas e lianas), para síndromes de dispersão seguimos Van der Pijl (1982) (anemocoria, zoocoria e autocoria) e para indicadores do estágio de regeneração, classificamos apenas as espécies arbóreas, conforme Gandolfi *et al.* (1995) (pioneira, secundária inicial e secundária tardia).

## 2.3 Análise dos dados

Todas as análises foram realizadas no programa R, versão 3.3.3, e os códigos da rotina analítica empregada estão disponíveis no Research Gate e pode ser baixada no link: [doi.org/10.13140/RG.2.2.31564.97920](https://doi.org/10.13140/RG.2.2.31564.97920)

Calculamos os parâmetros fitossociológicos de cada espécie para ambas as áreas (MUELLER-DOMBOIS & ELLENBERG, 1974). Utilizamos os índices de Diversidade de Shannon ( $H'$ ), o qual considera igual peso entre as espécies raras e abundantes (MAGURRAN, 2004) e o índice de equabilidade de Pielou ( $J'$ ), derivado do índice de diversidade de Shannon que permite representar a uniformidade da distribuição dos indivíduos entre as espécies existentes (PIELOU, 1966).

Para comparação dos valores médios de riqueza, abundância e diversidade de espécie das parcelas entre as duas áreas independentes utilizamos o teste  $t$ . Os pressupostos de distribuição normal e homogeneidade das variâncias foram verificados usando o testes Shapiro-Wilk e Levene, respectivamente. Para riqueza obtivemos distribuição normal e homogeneidade das variâncias, para abundância obtivemos a distribuição normal (dados logaritmizados) e heterogeneidade das variâncias e para diversidade distribuição normal e heterogeneidade das variâncias.

Para comparar a relação de riqueza, abundância e equitabilidade das duas áreas, construímos um gráfico de rank-abundância para complementar as interpretações sobre distribuição de composição de espécies e número de indivíduos.

Avaliamos a similaridade da composição de espécies e da estrutura de indivíduos entre os BSS das duas áreas, utilizando o Escalonamento Multidimensional Não-métrico (NMDS). Para as duas análises consideramos, respectivamente, a matriz de presença e ausência de espécies e o Índice de Jaccard como medida de similaridade florística e a matriz de abundâncias dos indivíduos e o Bray Curtis como índice de similaridade.



### 3 | RESULTADOS

A lista de espécies e a tabela dos parâmetros fitossociológicos estão disponíveis para consulta no Research Gate e pode ser baixada no link: [doi.org/10.13140/RG.2.2.23176.37124](https://doi.org/10.13140/RG.2.2.23176.37124)

Contabilizamos na área queimada 1.127 plântulas do BSS, distribuídas em 11 famílias, 16 gêneros e 26 espécies. Na área referência foram 683 plântulas do BSS, distribuídas em nove famílias, 11 gêneros e 27 espécies.

A densidade da área queimada foi de 375,6 indivíduos / m<sup>2</sup> e na área referência foi de 227,6 indivíduos / m<sup>2</sup>. Na a área queimada quatro espécies correspondem à 82,07 % de todos indivíduos, sendo três espécies de ervas e uma arbustiva. Suas respectivas densidades foram, *Scoparia dulcis* (129 indivíduos / m<sup>2</sup>), *Cyperus aggregatus* (85,3 indivíduos / m<sup>2</sup>), *Brachiaria plantaginea* (60,6 indivíduos / m<sup>2</sup>) e a arbustiva *Phytolacca thyrsoiflora* (33,33 indivíduos / m<sup>2</sup>). Na área referência as espécies que apresentaram maiores valores de densidade foram as ervas *S. dulcis* (129 indivíduos / m<sup>2</sup>) e Poaceae 3 (16 indivíduos / m<sup>2</sup>) e as arbustivas Indeterminada 3 (13,33 indivíduos / m<sup>2</sup>) e *Leandra* 2 (13 indivíduos / m<sup>2</sup>), correspondendo a 75,25 % de todos indivíduos.

Na área queimada as espécies com maiores Índice de Valor de Importância (IVI) foram *S. dulcis*, *C. aggregatus* e *B. plantaginea* e na área referência foram *S. dulcis*, Poaceae 3 e *Leandra* 2.

Os índices de diversidade de Shannon e de equabilidade de Pielou, respectivamente, foram  $H' = 1,973$  e  $J' = 0,605$  para a área queimada e  $H' = 1,825$  e  $J' = 0,598$  para a área referência.

Os valores médios de riqueza de espécies, abundância de indivíduos e diversidade de espécies das parcelas foram maiores na área queimada em relação aos da área referência (Figura 2).



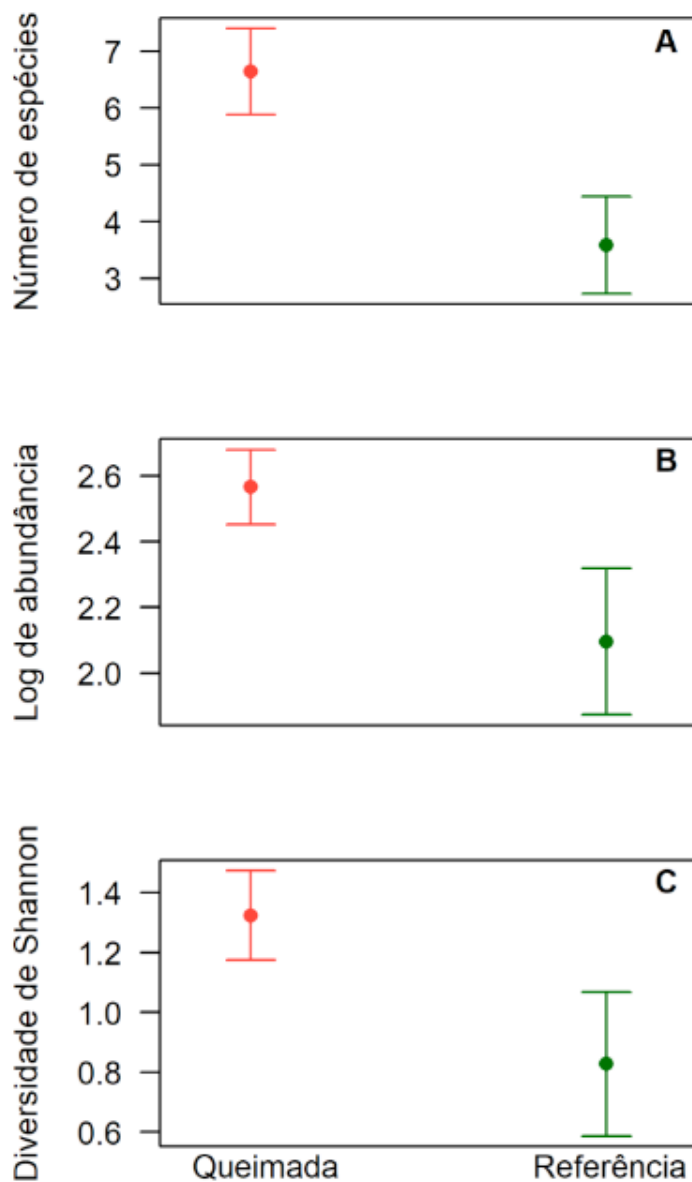


Figura 2. Valores médios para riqueza de espécies, abundância de indivíduos e diversidade de espécies das parcelas amostradas nas áreas queimada (vermelho) e referência (verde) no MNEI.

Nos gráficos de rank-abundância vimos que a área queimada possui maior abundância de indivíduos, porém a área referência apresenta maior equabilidade, apresentando menor variação na distribuição das abundâncias entre as espécies, comparada à área queimada (Figura 3).

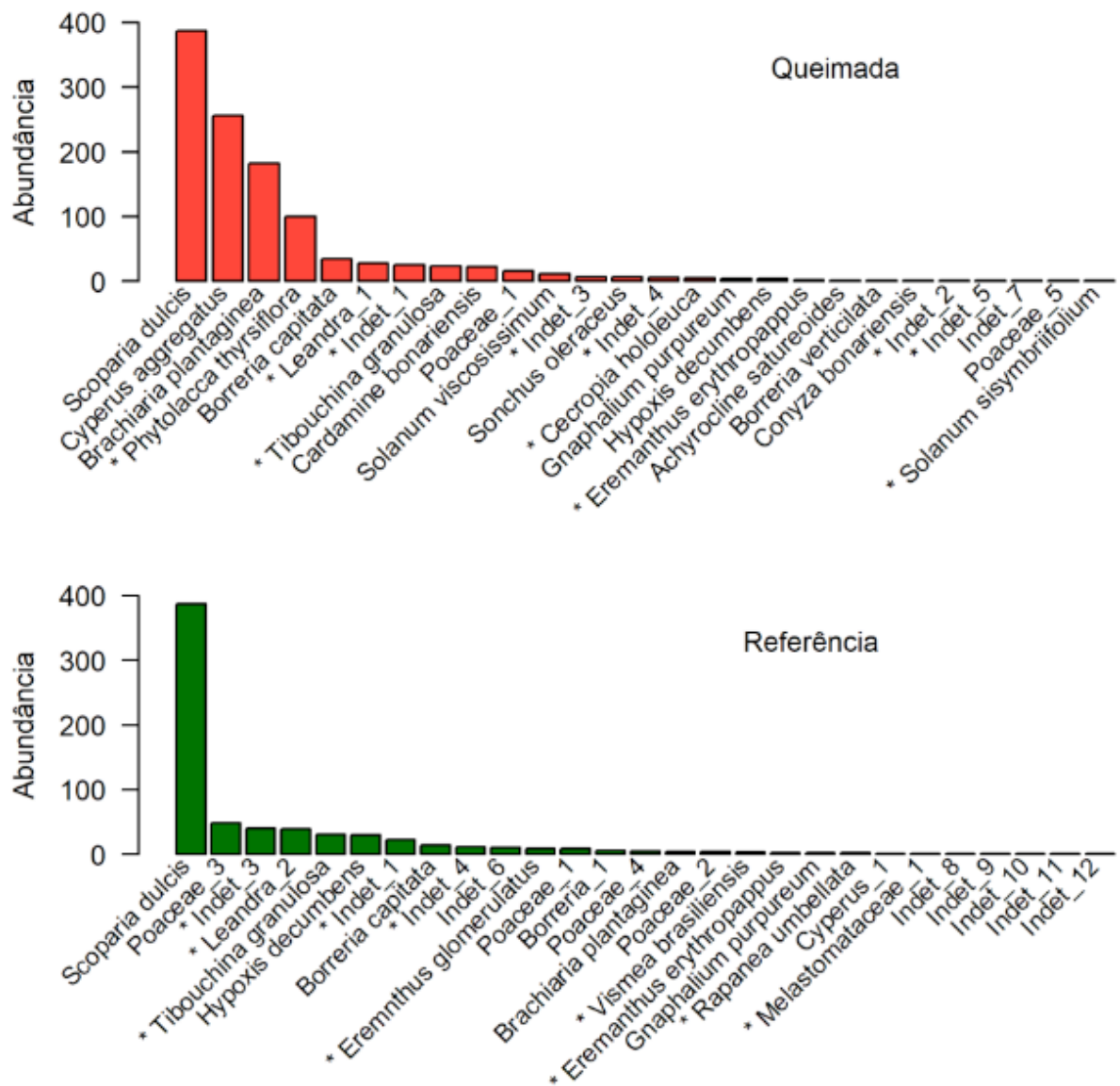


Figura 3. Gráfico de Rank-abundância para as áreas queimada (vermelho) e referência (verde) no MNEI.

Na análise de NMDS (Figura 4) vimos a separação dos ambientes área queimada e área referência, tanto para os dados de composição (riqueza) quanto para os de estrutura (abundância), com os valores de stress 0,125 e 0,135, respectivamente. Percebemos que a área referência apresenta maior variabilidade na composição de espécies e na estrutura dos indivíduos, em relação a área queimada. Os vetores arbustiva ( $p = 0,008$ ) e erva ( $p = 0,016$ ) tiveram influência significativa na composição e o vetor erva ( $p = 0,016$ ) teve influência significativa na estrutura.

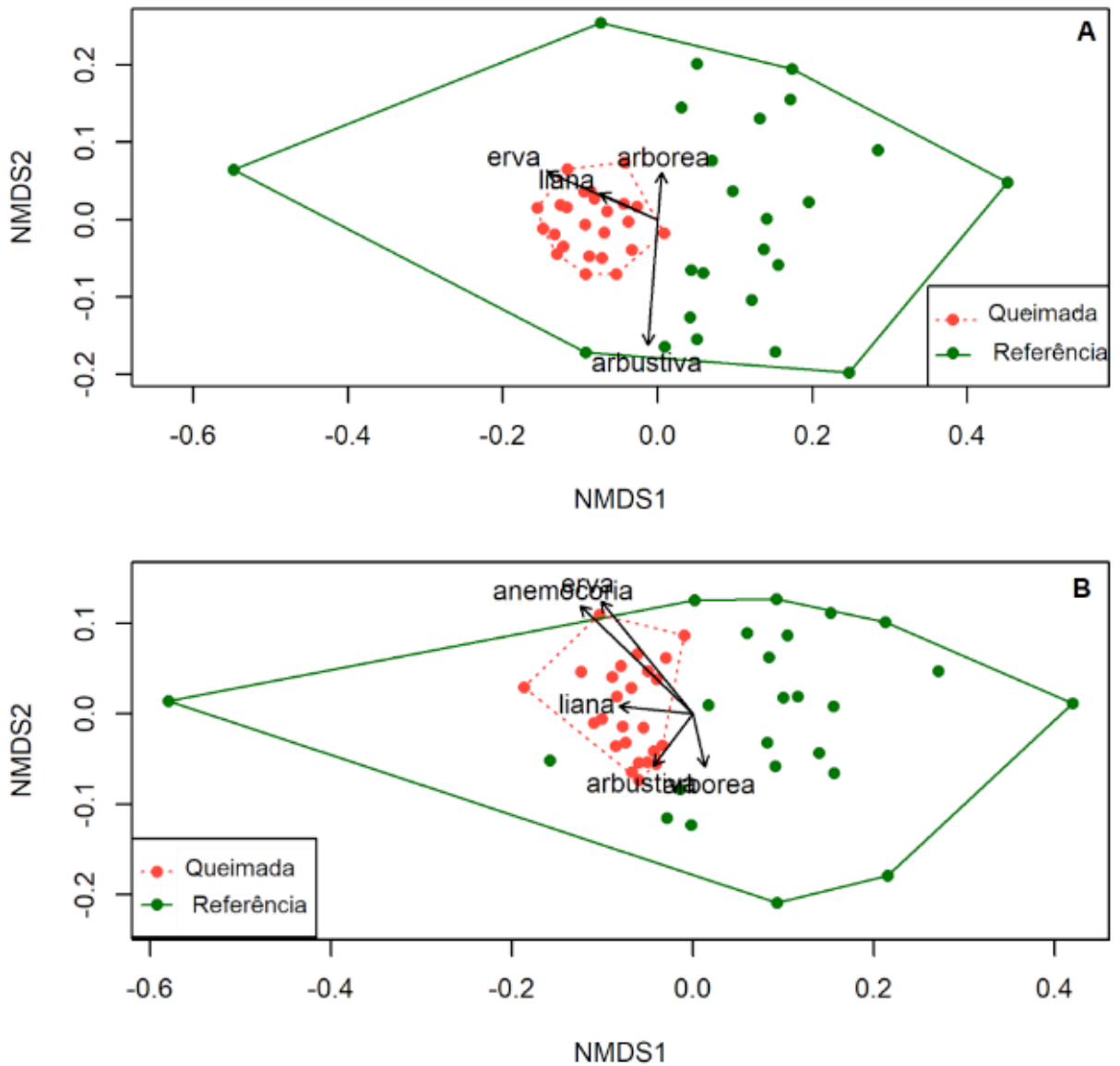


Figura 4. Similaridade para riqueza de espécies (A) e abundância de indivíduos (B) entre as áreas queimada (vermelho) e referência (verde) no MNEI.

Em relação as formas de vida, na área queimada, encontramos três espécies arbóreas (30 indivíduos), sete espécies arbustivas (168 indivíduos), 13 espécies de ervas (916 indivíduos) e uma espécie de liana. Na área referência encontramos cinco espécies arbóreas (47 indivíduos), cinco espécies arbustivas (113 indivíduos) e 13 espécies (521 indivíduos) de ervas.

Para a síndrome de dispersão encontramos na área queimada cinco espécies zoocóricas (134 indivíduos) e 15 espécies anemocóricas (941 indivíduos) e na área referência foram quatro espécies zoocóricas (45 indivíduos) e 14 espécies anemocóricas (552 indivíduos). Quanto as indicadoras do estágio de regeneração, na área queimada as três espécies arbóreas são pioneiras (30 indivíduos) e na área referência três espécies arbóreas são pioneiras (42 indivíduos) e duas espécies arbóreas são secundárias iniciais (5 indivíduos).

## 4 | DISCUSSÃO

Apesar da severidade do fogo, observamos que as áreas queimada e referência apresentaram valores similares de riqueza (26 e 27 espécies, respectivamente). Esses valores são inferiores aos encontrados por Melo & Duringan, (2007), 26 e 40 espécies respectivamente, e por Camargos *et al.* (2013), 15 e 16 espécies, respectivamente. Em nosso estudo a densidade de indivíduos foi maior na área queimada do que na área referência, oposto do encontrado nos dois estudos referidos acima. A maior densidade na área queimada foi devido à dominância de três espécies herbáceas competidoras (*S. dulcis*, *B. plantaginea* e *C. aggregatus*) e da arbustiva *P. americana* que corresponderam a 82,07 % dos indivíduos totais. Na área referência *S. dulcis* também esteve presente e em elevada densidade. A presença de herbáceas invasoras podem reduzir o avanço da trajetória sucessional sendo um filtro biótico que limita o estabelecimento de outras espécies, além de contribuir para potencialização de novos distúrbios pelo fogo devido à alta flamabilidade da biomassa de herbáceas (HOLL & AIDE, 2011).

Embora ambas as áreas avaliadas estejam inseridas em um grande contínuo florestal, que é fonte de sementes e animais dispersores, a paisagem é vulnerável à incidência do fogo devido aos encraves entre floresta e campo. BOND & KEELEY (2005) realizaram um estudo comparando o fogo como consumidor da vegetação análogo aos herbívoros, e assim, sendo uma força evolutiva significativa na estruturação e distribuição dos ecossistemas. De acordo com esses autores um único evento de fogo pode levar a redução de 1 a 2/3 na riqueza de espécies arbóreas e dependendo da frequência das queimadas promovem a substituição dessas por espécies herbáceas, processo conhecido como savanização das florestas.

Comparando ambas as áreas queimada e referência vimos que a área queimada apresentou valores médios de riqueza florística, densidade de indivíduos e Índice de diversidade de Shannon por parcela, maiores que a área referência. Esses resultados isoladamente nos levam a interpretação de que área queimada apresenta maior resiliência em relação a referência. Porém, esse resultado é devido a elevada densidade das quatro espécies invasoras dominantes (três herbáceas e uma arbustiva) da área queimada, mencionadas acima. A análise de NMDS também corrobora essa explicação ao mostrar que área queimada apresenta menor variabilidade, tanto para a composição de espécies quanto para a estrutura de indivíduos por parcela.

Em nossos resultados ambas as áreas não diferiram muito em relação a composição dos grupos ecológicos. Quanto a forma de vida tivemos maior riqueza e abundância para ervas, arbustivas e arbóreas, respectivamente. Áreas em que o uso da terra para pecuária e/ou queimada ocorre há vários anos tem o banco

de sementes severamente comprometido na composição de espécies arbóreas (CHAZDON, 2012).

A síndrome de dispersão mais abundante foi a anemocoria em ambas as áreas. Comparando a zoocoria entre as duas áreas percebemos que a elevada abundância da área queimada foi devido aos 100 indivíduos da espécie arbustiva invasora *P. americana*. Baixas taxas de espécies zoocóricas em florestas sugerem a perturbação desses ambientes (PENHALBER & MANTOVANI, 19970). Espécies arbóreas e zoocóricas, como *Trema micranta* e *Cecropia* sp., importantes na regeneração de clareiras e distúrbios antrópicos (BRAGA *et al.*, 2016) não foram encontradas na área referência e apenas um indivíduo de *C. hololeuca* esteve presente na área queimada.

Quanto a classificação das arbóreas indicadoras do estágio de regeneração natural encontramos a dominância de espécies e indivíduos pioneiros, com apenas duas espécies secundárias iniciais (cinco indivíduos) ocorrentes na área referência. Em ambas as áreas a avaliação da composição, estrutura e diversidade do banco de sementes mostrou um comprometimento da resiliência atual dessas áreas, sobretudo a área afetada pelo fogo, que apresentou elevada densidade de espécies herbáceas invasoras. Os gráficos de rank-abundância para cada área ilustraram essas relações de riqueza, abundância e uniformidade. Como as espécies tem estratégias de crescimento diferentes, tais mudanças ambientais locais também levam à mudanças na composição de espécies ao longo do tempo. Por tanto, entender como a sucessão funciona é crucial para melhorar as práticas de restauração de florestas e para selecionar as espécies mais adequadas a serem plantadas (ROZENDAAL *et al.*, 2019; POORTER *et al.*, 2019).

## 5 | CONCLUSÕES E RECOMENDAÇÕES

A ocorrência de fogo frequente na vegetação florestal tende, com o tempo, a levar a um empobrecimento de espécies arbóreas no banco de sementes do solo e gerar um modelo de savanização, com aumento das áreas de campo e redução das áreas florestais. Por se tratarem de áreas inseridas em uma UC, técnicas de restauração ecológica que aliem a eficácia e baixos custos são indicadas por se enquadrarem melhor dentro dos orçamentos públicos. A técnica de plantio de mudas arbóreas, em pequenos núcleos de diversidade, espalhados pela área, pode acelerar o processo de regeneração natural, promovendo o sombreamento ao redor dos núcleos, o que inibe a competição das gramíneas e facilita a entrada de outras espécies tolerantes a sombra (MARTINS, 2013). Indicamos o plantio de arbóreas zoocóricas para atrair a fauna dispersora de sementes e potencializar o processo

de formação do banco de sementes e o plantio de espécies pioneiras, secundárias iniciais e tardias, para promover a substituição de espécies no processo de sucessão ecológica.

## AGRADECIMENTOS

À CAPES pela concessão da bolsa durante os quatro anos de pesquisa. Ao Instituto Estadual de Florestas de Minas Gerais, órgão gestor das duas UCs nas quais pesquisamos (MNEI e PESOB). Agradecemos toda a disponibilidade de informações e logística em campo. Aos estudantes de pós-graduação do Laboratório de Restauração Florestal da UFV, Camila Bauchspiess e Lucas Siqueira Cardinelli e aos estagiários Leonardo Guimarães Ziccardi e Tales Thomaz Veloso, pelo auxílio em campo nas coletas do banco de sementes.

## REFERÊNCIAS

BOND, W. J.; KEELEY, J. E. **Fire as a global 'herbivore': the ecology and evolution of flammable ecosystems.** Trends in Ecology and Evolution, v. 20, n. 7, p. 387-394, 2005.

BRAGA, A. J. T.; BORGES, E. E. L.; MARTINS, S. V. **Seed bank in two sites of semideciduous seasonal forest in Viçosa, Minas Gerais.** Revista Árvore, v. 40, n. 3, p. 415-425, 2016.

CAMARGOS, V. L.; MARTINS, S. V.; RIBEIRO, G. A.; CARMO, F. M. S.; SILVA, A. F. **Influência do fogo no banco de sementes do solo em floresta estacional semidecidual.** Ciência Florestal, v.23, n. 1, p. 19-28, 2013.

CHAZDON, R. **Regeneração de florestas tropicais.** Boletim do Museu Paraense Emílio Goeldi - Ciências Naturais, v. 7, n. 3, p. 195-218, 2012.

GANDOLFI, S.; LEITÃO FILHO, H. F.; BEZERRA, C. L. **Levantamento florístico e caráter sucessional das espécies arbustivo-arbóreas de uma floresta mesófila semidecidual no município de Guarulhos, SP.** Revista brasileira de biologia, v. 55, n. 4, p.753-767,1995.

HARDESTY, J.; MYERS, R.; FULKS, W. **Fire, ecosystems and people: A preliminary assessment of fire as a global conservation issue.** Fire Management, v. 22, n. 4, p. 78-87, 2005.

HARPER, J. L. **Population biology of plants.** London: Academic Press, 1977. 892 p.

HENRIQUES, R. P. B. **Influência da história, solo e fogo na distribuição e dinâmica das fitofisionomias do bioma Cerrado.** In: SCARIOT, A.; SOUSA-SILVA, J. C.; FELFILI, J. M. Brasília: MMA, 2005. p. 73-92.

HERBÁRIO VIRTUAL REFLORA - **Flora do Brasil 2020 em construção.** Jardim Botânico do Rio de Janeiro. Disponível em: < <http://floradobrasil.jbrj.gov.br/> >. Acesso em: 20 ago. 2017.

HOLL, K. D.; AIDE, T. M. **When and where to actively to restore ecosystems?** Forest Ecology and Management, v. 261, p. 1558-1563, 2011.

HOWE, H. F.; SMALLWOOD, J. **Ecology of seed dispersal.** Annual Review of Ecology and Systematics, v. 13, n. 1, p. 201-228, 1982.



- MAGURRAN, A. E. **Measuring biological diversity**. Oxford: Blackwell Science, 2004. 256p.
- MARTINS, S. V. **Recuperação de áreas degradadas: ações em áreas de preservação permanente, voçorocas, taludes rodoviários e de mineração**. Viçosa: Aprenda Fácil, 2013.
- MARTINS, S. V.; RODRIGUES, R. R.; GANDOLFI, S.; CALEGARI, L. In: MARTINS, S. V. **Ecologia de florestas tropicais do Brasil**. 3. ed. Viçosa: UFV, 202. p. 21-52.
- MELO, A. C. G.; DURINGAN, G.; GORENSTEIN, M. R. **Efeito do fogo sobre o banco de sementes em faixa de borda de Floresta Estacional Semidecidual, SP, Brasil**. Acta Botanica Brasilica, v. 21, n. 4, p. 927-934, 2007.
- MINAS GERAIS. **Nota técnica para criação do Parque Estadual Serra do Ouro Branco e Monumento Natural Estadual de Itatiaia**. Instituto Estadual de Florestas, Governo de Minas Gerais, Belo Horizonte, MG, 2009.
- MIRANDA, H. S.; BUSTAMANTE, M. M. C.; MIRANDA, A. C. **The fire factor**. In: OLIVEIRA, P. S.; MARQUIS, R. J. *The cerrados of Brazil: Ecology and natural history of a neotropical savanna*. New York: Columbia University Press, 2002. p. 51-68.
- MUELLER-DOMBOIS, D.; ELLENBERG, H. **Aims and methods of vegetation ecology**. New York: John Wiley; Sons, 1974.
- PENHALBER, E. F.; MANTOVANI, W. **Floração e chuva de sementes em mata secundária em São Paulo, SP**. Revista Brasileira de Botânica, v. 20, p. 205–220, 1997.
- PIELOU, E. C. **The measurement of diversity in different types of biological collections**. Journal of Theoretical Biology, v. 13, p. 131 – 44, 1966.
- POORTER, L.; ROZENDAAL, D. M.; BONGERS, F.; et al. **Wet and dry tropical forests show opposite successional pathways in wood density but converge over time**. Nature Ecology and Evolution, v. 3, p. 928-934, 2019.
- ROZENDAAL, D. M. A.; BONGERS, F.; AIDE, T. M.; ALVAREZ-DÁVILA, E.; et al. **Biodiversity recovery of Neotropical secondary forests**. Science Advances, v. 5, n. 3, p. 1-10, 2019.
- VAN DER PIJL, L. **Principles of dispersal in higher plants**. 3. ed. Berlim: Springer-Verlag, 1982.

## EFFECTS OF SPATIAL SCALE ON PEQUI ENTOMOFAUNA

Data de aceite: 12/02/2020

Data de submissão: 28/11/2019

### Gustavo Amorim Santos

Universidade Federal de Goiás, Instituto de  
Química  
Goiânia – Goiás

[http://buscatextual.cnpq.br/buscatextual/  
visualizacv.do?id=K4210157E7](http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4210157E7)

### Deomar Plácido da Costa

Instituto Federal Goiano,  
Ceres – Goiás

[http://buscatextual.cnpq.br/buscatextual/  
visualizacv.do?id=K4744387Y9](http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4744387Y9)

### Suzana da Costa Santos

Universidade Federal de Goiás, Instituto de  
Química  
Goiânia – Goiás

[http://buscatextual.cnpq.br/buscatextual/  
visualizacv.do?id=K4793945U9](http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4793945U9)

### Pedro Henrique Ferri

Universidade Federal de Goiás, Instituto de  
Química  
Goiânia – Goiás

[http://buscatextual.cnpq.br/buscatextual/  
visualizacv.do?id=K4785526A5](http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4785526A5)

**ABSTRACT:** Spatial distribution and abundance of entomofauna can alter plant community structure by affecting the

competitive ability of individual plants. Spatial influence on the diversity of leaf entomofauna in eleven populations of *Caryocar brasiliense* was determined at large, intermediate and small scales using Moran's eigenvector maps (MEMs) and multivariate variograms. The insect community was best captured by MEMs derived from the Gabriel graph which was linearly weighted in terms of distance between sampling sites. Spatial dependencies were measured in the phenolic contents of pequi leaves and soil nutrients, which suggests that insect diversity is spatially dependent on these variables. Large and intermediate spatial scales were mainly related to the highest  $Mn^{2+}$  and  $Cu^{2+}$  soil contents and to the lowest total phenol contents of pequi leaves. MEMs not only enabled the modelling of multiscale spatial influence in the study area, but also controlled the spatial variance not explained by phenolic contents and soil nutrients.

**KEYWORDS:** *Caryocar brasiliense*, insects, chemical variability, soil nutrients, spatial variation.

### EFEITOS DA ESCALA ESPACIAL NA ENTOMOFAUNA DO PEQUI

**RESUMO:** A distribuição espacial e abundância da entomofauna podem alterar a estrutura de uma comunidade de plantas, afetando a

capacidade competitiva de plantas individuais. A influência espacial na diversidade da entomofauna foliar em onze populações de *Caryocar brasiliense* foi determinada em escalas grande, intermediária e pequena, utilizando mapas de autovetores de Moran (MEMs) e variogramas multivariados. A comunidade de insetos foi melhor representada pelos MEMs derivados do grafo de Gabriel ponderado linearmente com a distância entre os locais de amostragem. Dependências espaciais foram observadas quanto aos teores fenólicos das folhas de pequi e nutricionais do solo, sugerindo que a diversidade de insetos é espacialmente dependente dessas variáveis. As estruturas espaciais de escalas grande e intermediária foram relacionadas principalmente aos maiores teores de  $Mn^{2+}$  e  $Cu^{2+}$  do solo e ao menor teor de fenóis totais das folhas. Os MEMs não apenas permitiram modelar a influência espacial em múltiplas escalas na área de estudo, como também controlaram a variação espacial não explicada pelos conteúdos fenólicos e nutrientes do solo.

**PALAVRAS-CHAVE:** *Caryocar brasiliense*, insetos, variabilidade química, nutrientes do solo, variabilidade espacial.

## 1 | INTRODUCTION

Insect richness results from the balance of several ecological factors (ABDALA-ROBERTS et al., 2019). Insects, which amount to nearly 60% of all animals, are especially useful as bioindicators of environmental changes (ZEQUI et al., 2019). Climate, plant-insect interaction, habitat complexity, and host plant quality have generally been used to explain variations in insect diversity within and between habitats (TORRES & MADI-RAVAZZI, 2006). Plant metabolites can also affect entomofauna distribution and, to some extent, result from responses to the local abiotic and biotic environment (MUTIKAINEN et al. 2000), despite being genetically determined and potentially inducing the presence of herbivores (MEAUX & MITCHELL-OLDS, 2003). Soil nutrients can also affect plant metabolites (CUNNINGHAM et al., 1999), although the consequences for plant-insect interactions are still little explored (MEINDL et al., 2013).

Food resources and soil nutrients vary in space, directly affecting the diversity and distribution of insect populations (BISPO & OLIVEIRA, 2007). Although several studies have assessed spatial influence on insect and plant distribution, few have simultaneously analyzed entomofauna distribution with the host plant's metabolite variability and abiotic factors such as soil nutrients. The joint effects of these factors have recently been described on leaves of *Caryocar brasiliense* Cambess (Caryocaraceae), a species tree found in several phytophysiognomies of the Brazilian Cerrado and popularly referred as pequi.

The diversity of pequi entomofauna, when restricted to its leaves (FERREIRA et al., 2009), shows variation fractions better explained by edaphic (64.5%) and spatial

(60.6%) predictors than phenolic contents (22.7%) in the leaves (COSTA et al., 2019). The spatial pattern observed by Costa et al. used polynomial expressions of spatial coordinates as predictors, which cannot accurately model the spatial patterns on intermediate and small scales (JOMBART et al., 2009). Thus, empirically observed spatial influence may stem from a multiscale process, whose identification is a crucial issue in ecological studies (DUNGAN et al., 2002).

In this study, spatial influence on diversity of insects in pequi was partitioned into the spatial components of large, intermediate and small scales, by applying Moran's eigenvector maps (MEMs) as well as multiscale ordination (MSO) using multivariate variograms (WAGNER, 2004). Thus, the aim was to produce a more effective description of the associations between the entomofauna and phenolic contents of pequi leaves, in addition to the soil nutrients sampled in the study sites.

## 2 | MATERIAL AND METHODS

**Data Collection and Analysis:** Leaf entomofauna was collected from 11 pequi populations located in the cities of Alvorada do Norte (ALV), Araguapaz (ARA), Damianópolis (DAM), Faina (FAI), Formoso (FOR), Hidrolândia (HID), Iporá (IPO), Mambaí (MAM), Morrinhos (MOR), Mutunópolis (MUT), and Paraúna (PAR), all in the central Brazilian Cerrado, in Goiás state. Detailed information on leaf collection and phenolic analysis (total phenols, total tannins, flavonoids, condensed tannins and hydrolysable tannins), as well as on the collection and analysis of 17 physicochemical soil parameters, has been previously described (COSTA et al., 2019). Canonical correspondence analysis (CCA) was applied to the response matrix formed by occurrence (presence/absence) of leaf insects (41 sampling sites x 95 species), conditioned by edaphic (41 x 17 soil parameters) and leaf chemicals (41 x 5 phenolic contents) defined as explanatory variables (COSTA et al., 2019). Successive CCAs allowed the selection of significant soil ( $\text{Cu}^{2+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Zn}^{2+}$ , P) and leaf (total phenols, flavonoids, total tannins and hydrolysable tannins) variables. In all CCAs, the variance inflation factor ( $\text{VIF} < 4$ ) with Holm-Bonferroni correction was used to guide the selection of explanatory variables, avoiding multicollinearity in multivariate regressions. VIF values  $> 20$  are considered strongly multicollinear (LEPS & SMILAUER, 2007).

Moran's eigenvector maps (MEMs) were generated by means of a spatial weighting matrix (**W**) obtained by the Hadamard product of a connectivity matrix (**B**) with weighting matrices (**A**) (DRAY et al., 2006; DRAY et al., 2019). Matrix **B** was generated using three graph-based connection schemes (Gabriel = *gab*, relative neighbourhood = *rel* and minimum spanning tree = *mst*). Weighting matrices **A** were derived from three functions:  $f_{\text{bin}}$  = binary, based only on topology and reflecting a neutrality in the weight;  $f_{\text{lin}} = 1 - (d/d_{\text{max}})$ , linear, and  $f_{\text{con}} = 1 - (d/d_{\text{max}})^{0.5}$ , nonlinear, both weighted by

the inverse of the power of Euclidean distances ( $d$  and  $d_{max}$ ) between two sampling sites (BAUMAN et al., 2018). The selection of matrix  $\mathbf{W}$  among the nine candidate matrices and the subset of significant MEMs was conducted using the optimization method (*listw.candidates* function) available in the *adespatial* package (DRAY et al., 2019), considering only MEMs associated with positive eigenvalues.

Significant MEMs were incorporated into the explanatory matrix containing previously selected soil nutrients and leaf phenolic contents. They were submitted to stepwise selection by CCA, whose best model was based on the  $VIF < 8$  and the adjusted coefficient of determination ( $R^2_{adj}$ ), allowing the comparison of different models. Pure and overlapping variance fractions of predictors were assessed by total variance partitioning via partial CCA (pCCA). In pCCA, four subsets of predictors were considered as explanatory: environmental variables, incorporating chemical and edaphic variables (total phenols, total tannins, flavonoids, hydrolysable tannins,  $Cu^{2+}$ ,  $Fe^{3+}$ ,  $Mn^{2+}$ ,  $Zn^{2+}$ , and P) and spatial predictors representing the large (MEM1, MEM3), intermediate (MEM5, MEM6) and small (MEM8, MEM14) scales previously selected by the CCA.

Multiscale ordination (MSO) was used to model spatial structures (WAGNER, 2004), with semivariance multiple comparisons adjusted by the Bonferroni correction ( $k = 0.05/n$ , where  $n$  is the number of classes and  $k$  the adjusted significance), avoiding the increase of error type I. Envelopes describing the confidence limits around the mean semivariance were used to assess the independence of scale over distance. The number of classes ( $h = 11$ ) was determined by Sturges' rule, which avoids arbitrary inflation of spatial variation explained as a function of class size. All analyses were conducted using *vegan* (OKSANEN et al., 2017), *spdep* (BIVAND & PIRAS, 2015) and *adespatial* (DRAY et al., 2019) packages, in R version 3.4.4 (R CORE TEAM, 2018).

### 3 | RESULTS AND DISCUSSION

All spatial weighting matrix  $\mathbf{W}$  displayed high statistical power, regardless of connectivity matrix  $\mathbf{B}$  (*gb*, *rel*, *mst*) between the sampling sites; weighting matrices ( $\mathbf{A}$ ) were either binary (without weight) or weighted by functions that decreased linearly or nonlinearly with the distance between two sites (Table 1).

$\mathbf{W}^a$	$R^2_{adj}$ (global) <sup>b</sup>	$R^2_{adj}$ (subset) <sup>c</sup>	No. of MEMs	$P^d$
<i>gb</i> -binary	0.86660	0.85700	17	0.00089
<b><i>gb</i>-linear</b>	<b>0.90096</b>	<b>0.89929</b>	<b>17</b>	<b>0.00089</b>
<i>gb</i> -nonlinear	0.87552	0.86779	17	0.00089
<i>rel</i> -binary	0.85395	0.85133	18	0.00089
<i>rel</i> -linear	0.89861	0.89214	17	0.00089

<i>rel</i> -nonlinear	0.87579	0.87116	17	0.00089
<i>mst</i> -binary	0.85395	0.85133	18	0.00089
<i>mst</i> -linear	0.89861	0.89214	17	0.00089
<i>mst</i> -nonlinear	0.87579	0.87116	17	0.00089

Table 1. Spatial weight matrices ( $\mathbf{W} = \mathbf{B}^* \mathbf{A}$ ) with adjusted coefficients of determination to the global model and subset of MEMs.

<sup>a</sup>Spatial weight matrix via graphs (*gb* = Gabriel, *rel* = relative neighbourhood, *mst* = minimum spanning tree).

<sup>b</sup> $R^2_{adj}$  for the global model and <sup>c</sup>subset of MEMs. <sup>d</sup>Probability with Sidak correction.

Among the nine candidate  $\mathbf{W}$  matrices, the one that used the linear weighting function with the Gabriel graph captured the highest variability of pequi leaf entomofauna. Optimization of  $\mathbf{W}$  and selection of the MEM subset were based on maximising the  $R^2_{adj}$  value and controlling error type I rate by the Sidak correction,  $P_s = 1 - (1 - P)^m$ , where  $m$  = number of tests. The graphs used to represent the connectivity matrix are inclusive, so that all *mst* connections are contained in those of *rel*, which in turn are those of *gb*, with the number of connections increasing along these graphs (LEGENDRE & LEGENDRE, 2012). However, Table 1 shows the same representation for *mst* and *rel*, probably as a result of sites' irregular arrangement. Simulations indicated that the performance of MEMs under irregular sampling is highly dependent on the connectivity matrix and, consequently, on an appropriate neighbourhood relationship (BRIND'AMOUR et al., 2018). In other simulated scenarios, connectivity matrices based on a Delaunay triangulation performed poorly, a finding associated with connections at the edges of sampling sites by Delaunay criterion (BAUMAN et al., 2018). These edge effects artificially connect distant sites, hence misspecifying matrix  $\mathbf{W}$  and causing the observed underestimations. Due to the sample's considerable irregularity, we chose not to incorporate the Delaunay graph among connectivity matrices.

The analysis of CCA using previously selected leaf chemical and soil nutrient variables allowed the selection of six significant MEMs (MEM1, MEM3, MEM5, MEM6, MEM8 and MEM14), based on  $VIF < 8$  and  $R^2_{adj}$  values. CCA results show that 91.7% ( $R^2_{adj} = 86.7\%$ ,  $F = 18.3$ ,  $p = 0.001$ ) of total variation in entomofauna can be explained by all selected predictors (total phenols, total tannins, flavonoids, hydrolysable tannins,  $\text{Cu}^{2+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Zn}^{2+}$ , P, MEM1, MEM3, MEM5, MEM6, MEM8 and MEM14) which showed strong correlations ( $R_1 = 0.962$ ;  $R_2 = 0.995$ ). Modelling took into account the occurrence (presence/absence) of entomofauna species, which seems to better reflect the community compared to the species richness index or distance-based approaches (LEGENDRE & LEGENDRE, 2012). The latter may overlap dispersion and local effects, leading to a failure to detecting a multivariate effect in ordination techniques (WARTON et al., 2012).

CCA1 (17.7%,  $F = 4.8$ ,  $p = 0.001$ ) was associated mainly with an increase in total



phenols, tannins and hydrolysable tannins, as well as with a reduction in flavonoid contents (Figure 1). These trends were observed in MEM5 and MEM3, respectively, which suggests that CCA1 is linked to metabolite variability. Moreover, CCA2 (14.8%,  $F = 4.8$ ,  $p = 0.001$ ) was mainly associated with a decrease in soil micronutrients mainly correlated to MEM1. CCA also show greater insect diversity in pequi populations located in micronutrient-rich soils (COSTA et al., 2019), such as those in Morrinhos (MOR), Araguapaz (ARA), Formosa (FOR), and Mutunópolis (MUT).

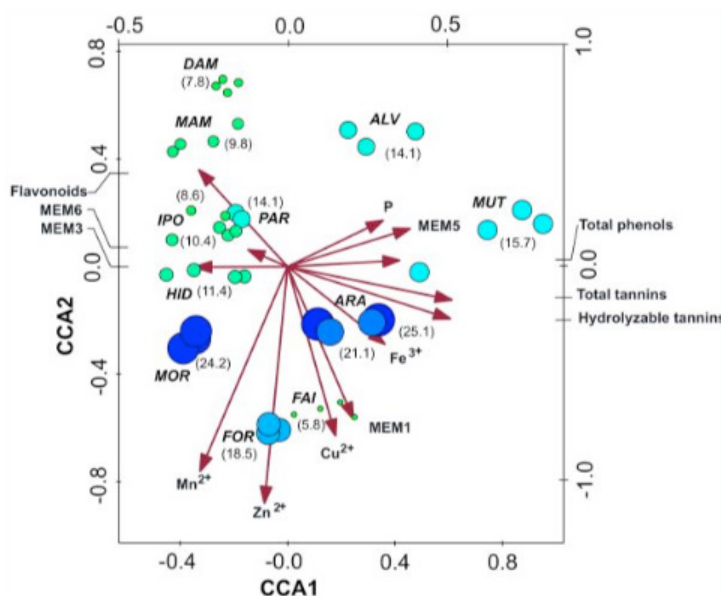


Figure 1. CCA triplot of pequi entomofauna with explanatory variables (red arrows). Pequi populations are represented by circles whose areas are proportional to Hill's measured diversity index (values in parentheses). MEM8 and MEM14 are not shown

Partitioning of the explained the variation provided  $R^2_{adj}$  values in pure fractions ( $p < 0.05$ ), without overlap of predictor effects, of 1.8% (leaf chemical), 43.8% (edaphic) and 33.2% (spatial), which were similar to 1.4%, 37.0% and 33.0%, respectively, obtained with polynomial functions of geographic coordinates (COSTA et al., 2019). MEMs represent a spectral decomposition of spatial relationships between observations in all possible variation ranges; thus, they represent different spatial scales. Those obtained by CCA adequately modelled spatial influence in large (MEM1, MEM3), intermediate (MEM5, MEM6) and small (MEM8, MEM14) scales (Figure 2).

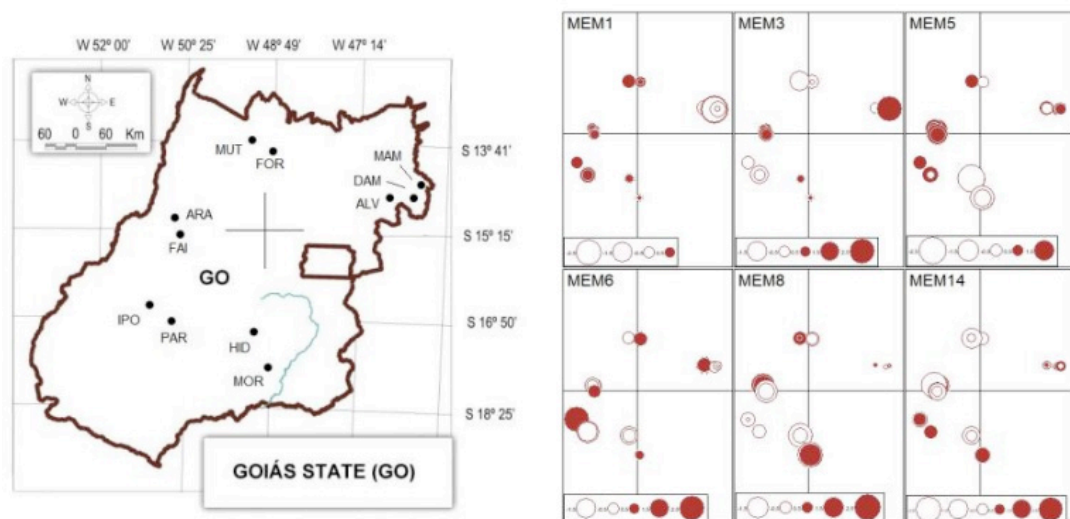


Figure 2. Pequi sampling sites and MEMs that modelled the large (MEM1, MEM3), intermediate (MEM5, MEM6) and small (MEM8, MEM14) scales of leaf entomofauna. MEMs with high positive values have larger brown bubbles.

Variance partitioning of MEMs by pCCA of leaf entomofauna indicated a larger  $R^2_{adj}$  for the intermediate scale (15.2%), followed by large (11.5%) and small scales (8.0%), which suggests pure spatial influences ( $p \leq 0.03$ ). Successive CCAs for each spatial scale resulted in canonical axes representing the three spatially structured patterns of entomofauna variation. Subsequent regression of site scores, viewed by significant canonical axes, on phenolic contents and soil nutrients made it possible to assess environmental influence on spatial patterns (Table 2).

Scale (MEMs)	Variable	Estimate	Standard error	t-Values	P (>  t  )
Large (MEM1, MEM3)	Fe <sup>3+</sup>	3.5886	0.9195	3.903	0.0005
	Mn <sup>2+</sup>	1.0241	0.3733	2.743	0.0100
	Total phenols	-3.4196	1.5186	-2.252	0.0316
Intermediate (MEM5, MEM6)	Cu <sup>2+</sup>	8.3212	1.6331	5.095	1.6e-05
	Fe <sup>3+</sup>	-1.9963	0.7908	-2.524	0.0169
	Total phenols	-2.9849	1.4805	-2.016	0.0525

Table 2. Regression parameters of spatial variation of significant canonical axes from leaf entomofauna with environmental variables.

As shown in Table 2, the large-scale spatial pattern of leaf entomofauna is clearly related to the highest Fe<sup>3+</sup> and Mn<sup>2+</sup> contents in the soil and the lowest total phenol contents in pequi leaves. The space-environment relationship observed at the intermediate scale, in turn, indicates an association with the highest and lowest levels of Cu<sup>2+</sup> and Fe<sup>3+</sup>, respectively. The absence of total phenols had only a weak influence in this scale. As it commonly observed, the structured small-scale spatial pattern had no relation with the environment. When significant, small-scale MEMs stem mainly from local spatial correlations generated within the community (LEGENDRE & LEGENDRE,

2012). These results are useful in explaining higher insect richness in pequi populations located in micronutrient-rich soils, as well as high phenolic content correlated with higher insect richness of Hymenoptera and Hemiptera in some populations, whereas lower values are mainly related to sites with a large number of beetles (COSTA et al., 2019).

Quantification of the influence of each subset in the space-environment relationship was obtained by pCCA, and results can be seen in the Venn diagram in Figure 3, whose values correspond to those of  $R^2_{adj}$ .

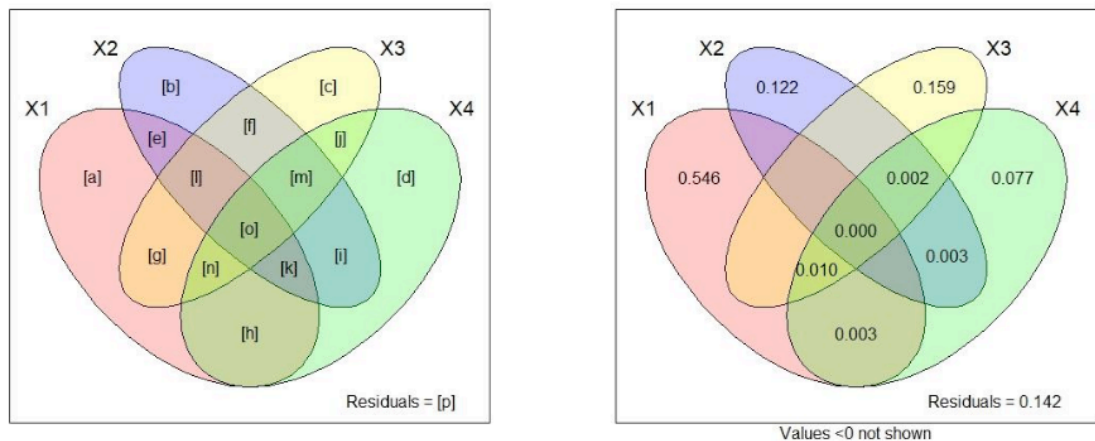


Figure 3. Partitioning of the explained variation of pequi leaf entomofauna by environmental descriptors (X1) (leaf phenolic contents and soil nutrients) and MEMs for large (X2), intermediate (X3) and small (X4) spatial scales. Negative values of  $R^2_{adj}$  are not shown.

Results show that 91.1% ( $R^2_{adj} = 85.8\%$ ) of the entomofauna variation ([a-o]) was explained by predictive matrices, resulting in a model with a residual of 14.2% ([h]). It was observed that 54.6% ([a]) of insect variability can be explained solely by soil characteristics and leaf phenolic contents. Pure spatial variation contributed with 35.8% ([b + c + d]), with major contributions to intermediate ([c] = 15.9%) and large ([b] = 12.2%) scales, and all pure fractions were regarded as significant ( $p = 0.001$ ). Partitioning generally assumes only additivity in predictor variables and can produce negative components of variation due to interactions between variables. In fact, spatial scale-related MEMs presented a nonzero intersection ([j + m + n + o] = 1.2%), despite being orthogonal. This is due to the fact that environmental variables involved in the pCCA are not orthogonal to MEMs, and the technique involves subtracting  $R^2_{adj}$  values which depend on the number of explanatory variables (LEGENDRE & LEGENDRE, 2012).

A solution to this problem was the combination of ordination techniques (CCA) with variograms, i.e. multiscale ordination (MSO). The MSO enables us to evaluate not only mean and variance stationarity in the study area, but also the absence of spatial correlation in regression residuals, hence allowing regression coefficients to be interpreted, since residuals are spatially independent (WAGNER, 2004).

Initially, total variation in distance classes ( $h$ ) was obtained without conditioning, that is, via correspondence analysis (CA) of the response matrix (entomofauna), followed by eigenvalue decomposition over distance (Figure 4A). Each point in the multivariate variogram represented the sum of all contributions of environmental variables to spatial covariance, weighted by sample pairs within the distance class and plotted as the mean of distance between the pairs. Total variance showed a sharp increase with distance, and spatial autocorrelation in two classes was statistically significant, which challenged the assumption of stationarity along distance  $h$ .

The entomofauna matrix was then conditioned by leaf phenolic contents and soil nutrients using CCA, whose eigenvalues were also decomposed along the distance classes (Figure 4B).

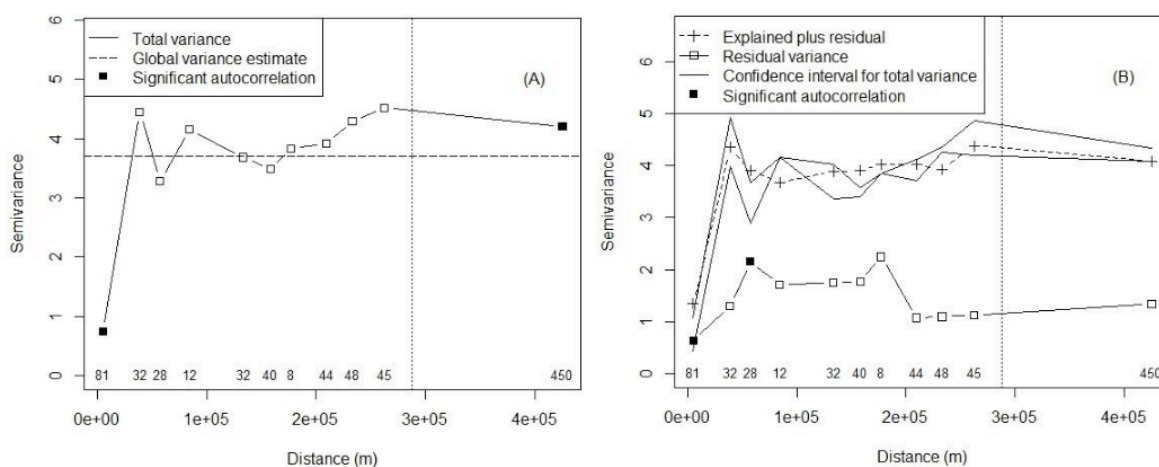


Figure 4. Multivariate variogram of (A) total variation of pequi entomofauna and (B) explained variation by environmental variables (leaf phenolic contents and soil nutrients).

Residual variance showed spatial autocorrelation in the first (distance < 5.1 km) and third (57.4 km) distance classes. CCA residues were spatially independent between and above these classes. Furthermore, when correlations between CCA matrices do not vary with the scale, the sum of explained and residual variances remains within a 5% probability envelope. In Figure 4B, only four classes remained within this confidence interval, and environmental predictors cannot be assumed to be implicitly invariant over a geographic scale.

In an attempt to control the regionalization of environmental variables caused by spatial autocorrelation, MSO was performed by introducing MEMs into the set of predictor variables (Figure 5A). Results show that the model controlled spatial autocorrelation in the residuals at distances of up to 60 km. This control can be attributed to the variation fraction captured by pCCA (fractions  $[n + h]$ , Figure 3), suggesting an environmental effect on the small and intermediate spatial distances. This is important because genotypic variability among neighboring trees is often associated with a spatial structure, regardless of environmental conditions (COVELO et al., 2011). However,

it could result from phenotypic characteristics of the individuals, which are strongly associated with autocorrelation on a small scale (LEGENDRE & LEGENDRE, 2012). However, this does not seem to be the case in this study.

Despite spatial autocorrelation control in the residuals of the regression models, the sum of explained and residual variances exceeded the confidence envelope in the fourth (85 km) and seventh (177 km) distance classes (Figure 5A). These results, compared with those given by variation partitioning (Figure 3), allow this finding to be attributed to the intersection of the spatial structure at different scales, represented by fractions  $[i + m]$  ( $R^2_{adj} = 0.5\%$ ). Lastly, CCA was performed with the explanatory environmental variables, i.e. phenolic contents and soil nutrients, now controlled by spatial structure, using MEMs as covariates. Figure 5B shows MEMs controlling scale dependence, so that the model was suitable for the environmental predictor variables.

In the multivariate variogram there is no spatial autocorrelation in regression residuals, and the sum of explained and residual variances after elimination of the spatial structure represented by MEMs remained within the confidence interval of total variance, including the fourth and seventh distance classes. In addition, MEMs removed the trend of total variance gradient of leaf entomofauna, thus resulting in a globally oscillating multivariate empirical variogram. Therefore, MEMs were able to control spatial variation in pequi leaf entomofauna, which was not explained by the predictive variables represented by leaf phenolic contents and soil nutrients.

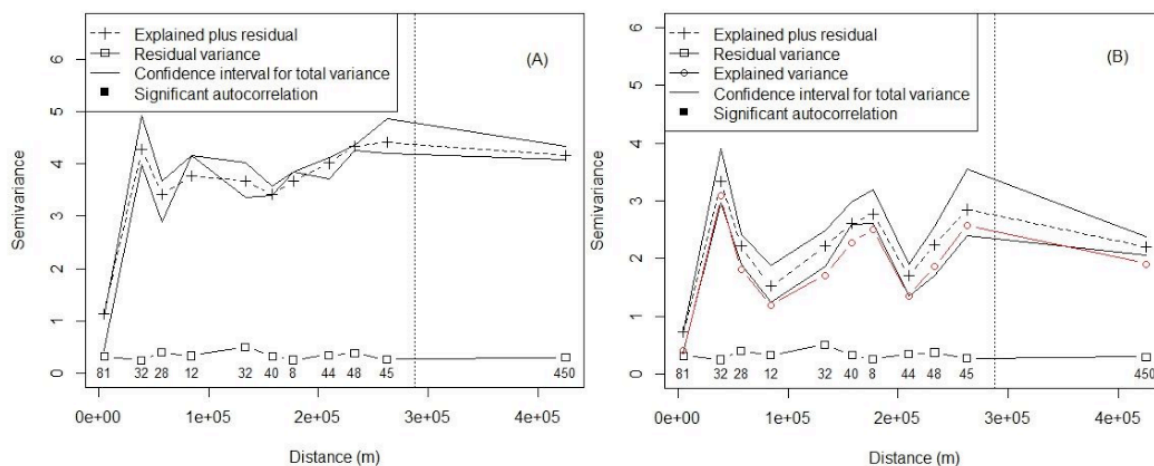


Figure 5. Multivariate variogram of pequi entomofauna (A) explained by environmental and spatial variables (MEMs) and (B) after removal of spatial structure explained by MEMs.

## 4 | CONCLUSIONS

Plant-soil interactions may influence entomofauna occurrence and diversity, despite being as yet largely unexplored. In this study, a significant spatial correlation was observed between pequi leaf entomofauna, on the one hand, and leaves' phenolic contents and soil mineral nutrients from sampling sites, on the other.



Given that several processes can create the same pattern of spatial structure, their identification alone is insufficient to elucidate the mechanisms underlying the spatial variability. In this study, MEMs not only enabled the modelling of spatial variability in the entomofauna, but also identified the associated spatial structure at large (MEM1 and MEM3), intermediate (MEM5 and MEM6) and small (MEM8 and MEM14) scales. Furthermore, multiscale ordination analysis showed that environmental influences, represented by leaves' chemical predictors and soil nutrients, are not only invariant but also congruent with the assumption of mean and variance stationarity between samples and of independence of residuals in multivariate regressions.

## ACKNOWLEDGEMENTS

The authors wish to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for its financial support (# 401188-2016-7).

## SUPPORTING INFORMATION

Additional supporting information (R script and dataset) may be found at <http://dx.doi.org/10.13140/RG.2.2.18335.92320>

## REFERENCES

- ABDALA-ROBERTS et al. Bottom-up control of geographic variation in insect herbivory on wild cotton (*Gossypium hirsutum*) by plant defenses and climate. **American Journal of Botany**, v. 106, n. 8, p. 1059-1067, 2019.
- BAUMAN, D. et al. Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. **Ecology**, v. 99, n. 10, p. 2159-2166, 2018.
- BISPO, P. C.; OLIVEIRA, L. G. Diversity and structure of Ephemeroptera, Plecoptera and Trichoptera (Insecta) assemblages from riffles in mountain streams of Central Brazil. **Revista Brasileira de Zoologia**, v. 24, n. 2, p. 283-293, 2007.
- BIVAND, R.; PIRAS, G. Comparing Implementations of Estimation Methods for Spatial Econometrics. **Journal of Statistical Software**, v. 63, n. 18, 1-36, 2015.
- BRIND'AMOUR, A. et al. Application of Moran's Eigenvector Maps (MEM) to irregular sampling designs. **Spatial Statistics**, v. 26, p. 56-68, 2018.
- COSTA, D. P. et al. Effects of soil, spatial parameters and leaf phenolic contents on entomofauna variability in pequi. In TULLIO, L. (Org.) **Características dos solos e sua interação com as plantas**. Atena: Ponta Grossa, Cap. 4, p. 28-42, 2019.
- COVELO, F. et al. Temporal changes in the spatial pattern of leaf traits in a *Quercus robur* population. **Annals of Forest Science**, v. 68, n. 3, p. 453-460, 2011.
- CUNNINGHAM, S. A. et al. Evolutionary divergences in leaf structure and chemistry, comparing rainfall and soil nutrient gradients. **Ecological Monographs**, v. 69, n. 4, p. 569-588, 1999.



- DUNGAN, J. et al. A balance view of scale in spatial analysis. **Ecography**, v. 25, n. 5, p. 626-640, 2002.
- DRAY, S. et al. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). **Ecological Modelling**, v. 196, n. 3/4, p. 483-493, 2006.
- DRAY, S. et al. **Adespatial: multivariate multiscale spatial analysis**. R package version 0.3-7, 2019. <https://CRAN.R-project.org/package=adespatial>.
- FERREIRA, G. A. et al. Biodiversidade de insetos em Pequiheiro (*Caryocar brasiliense*, Camb.) no Cerrado do Estado de Goiás, Brasil. **Agrociência**, v. 13, n. 2, p. 14-31, 2009.
- JOMBART, T. et al. Finding essential scales of spatial variation in ecological data: a multivariate approach. **Ecography**, v. 32, n. 1, p. 161-168, 2009.
- LEGENDRE, P.; LEGENDRE, L. **Numerical ecology**. 3th ed. Elsevier: Amsterdam, 2012.
- LEPS, J.; SMILAUER, P. **Multivariate analysis of ecological data using Canoco**. Cambridge University Press: Cambridge, 2007.
- MEAUX, J.; MITCHELL-OLDS, T. Evolution of plant resistance at the molecular level: ecological context of species interactions. **Heredity**, v. 91, n. 4, p. 345-352, 2003.
- MEINDL, G. A. et al. Edaphic factors and plant–insect interactions: direct and indirect effects of serpentine soil on florivores and pollinators. **Oecologia**, v. 173, n. 4, p. 1355-1366, 2013.
- MUTIKAINEN, P. et al. Herbivore resistance in *Betula pendula*: effect of fertilization, defoliation, and plant genotype. **Ecology**, v. 81, n. 1, p. 49-65, 2000.
- OKSANEN, J. et al **Vegan: community ecology package**. R version 2.4-5, 2017. <https://CRAN.R-project.org/package=vegan>.
- R CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing: Vienna, version 3.4.4, 2018.
- TORRES, F. R. & MADI-RAVAZZI, L. Seasonal variation in natural populations of *Drosophila* spp. (Diptera) in two woodlands in the State of São Paulo, Brazil. **Iheringia: Série Zoologia**, v. 96, n. 4, p. 437-444, 2006.
- WAGNER, H. H. Direct multi-scale ordination with canonical correspondence analysis. **Ecology**, v. 85, n. 2, p. 342-351, 2004.
- WARTON, D. I. et al. Distance-based multivariate analyses confound location and dispersion effects. **Methods in Ecology and Evolution**, v. 3, n. 1, p. 89-101, 2012.
- ZEQUI, J. A. C. et al. Aquatic insect communities in small stream in the south of Brazil. **Environmental Monitoring and Assessment**, v. 191, n. 7, p. 408, 2019.

## PIPELINE DE EXPRESSÃO DIFERENCIAL EM R APLICADO À *Arabidopsis thaliana*

Data de aceite: 12/02/2020

### Sheila Tiemi Nagamatsu

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Lucas Miguel de Carvalho

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Luciana Souto Mofatto

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Nicholas Vinícius Silva

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Marcelo Falsarella Carazzolle

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Gonçalo Amarante Guimarães Pereira

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

**RESUMO:** Devido a imensa complexidade genômica das plantas, a análise de transcriptômica assume um papel relevante para entendimento do mecanismo biológico, visto que além de dar indícios de como o organismo responde em determinada situação, o sequenciamento de RNA também pode ser utilizado para auxiliar a resolução de problemas de ordem genômica. A análise de expressão diferencial pode ser dividida em quatro etapas: i) análise de qualidade e trimagem dos dados; ii) alinhamento dos fragmentos à referência; iii) quantificação dos transcritos; e iv) análise de expressão diferencial. Em geral, apenas a etapa de análise de expressão diferencial (iv) é realizada em linguagem R, porém, nesse capítulo apresentamos uma versão diferenciada, em que os softwares externos também foram executados por comandos no R. Como estudo de caso foi utilizado um experimento de *Arabidopsis thaliana*, um organismo modelo que apresenta um ciclo de vida curto, é facilmente manipulado em laboratório e tem seu genoma bem estudado.

**PALAVRAS-CHAVE:** Transcriptômica de plantas, expressão diferencial, *Arabidopsis thaliana*, RNA-seq

**ABSTRACT:** Due to the immense complexity of plant genomes, transcriptomic analysis plays a relevant role in understanding the biological

mechanism, since in addition to providing evidence of how the organism responds in a given situation, RNA sequencing can also be used to aid genomic information. The analysis can be divided into four steps: i) quality analysis and data trimming; ii) alignment of fragments to the reference; iii) quantification of transcripts; and iv) differential expression analysis. In general, only the differential expression analysis step (iv) is performed in R language, but in this chapter we present a different version, in which external software was executed by commands in R. As an example was used an experiment of *Arabidopsis thaliana*, a model organism with a short life cycle, a well studied genome and easy to manipulate in laboratory.

**KEYWORDS:** Plant transcriptomics, differential expression, *Arabidopsis thaliana*, RNA-seq

## 1 | INTRODUÇÃO

Com o avanço da biologia molecular e das técnicas de sequenciamento foram desenvolvidas as ômicas; por exemplo, genômica, transcriptômica e proteômica. Cada uma dessas ômicas têm suas particularidades e nos auxiliam a responder perguntas específicas sobre o metabolismo celular. A genômica é utilizada para caracterizar um organismo a nível de DNA, molécula que é transmitida aos descendentes e que contém todo o potencial de genes que podem ser expressos em uma célula. A transcriptômica, por sua vez, surge para estudar o RNA das células, a parte do genoma que realmente é produzida em uma condição específica, sendo também utilizado para montagem dos transcritos (transcriptoma). A proteômica, por sua vez, visa estudar o conjunto de proteínas encontradas na célula em uma determinada situação.

Nesse contexto, apesar da transcriptômica ser utilizada principalmente para identificar expressão de genes diferenciais em condições específicas, ela também pode ser empregada para suprir demandas de análises genômicas, como, por exemplo, no auxílio da montagem e anotação de genomas em plantas complexas (Nascimento et al., 2019); busca de variantes genômicas (polimorfismos de nucleotídeos únicos, deleções e inserções de bases) (Chopra et al., 2015); e identificação de pequenas regiões repetitivas no genoma (microssatélites) em regiões codantes, utilizadas como marcadores genéticos (Bazzo, de Carvalho, Carazzolle, Pereira, & Colombo, 2018).

Dentre as metodologias *high-throughput* o RNA-seq recebe grande destaque. Esta metodologia apresenta como grande vantagem a possibilidade de identificação de novos genes, além de ser um método de maior acurácia quanto ao nível de quantificação gênica (Edwards & Batley, 2004; Hrdlickova, Toloue, & Tian, 2017). A primeira etapa para esse sequenciamento é a preparação das bibliotecas de cDNA, que deve ser específica para o tipo de RNA que será analisado. Ela é composta,

principalmente, pelas etapas: i) seleção de RNA de interesse na amostra celular; ii) depleção de RNA ribossomal em experimentos que não o tenham como objeto de estudo; iii) fragmentação do RNA; e iv) ligação de adaptador e identificador da biblioteca na extremidade dos fragmentos (Hrdlickova, Toloue, & Tian, 2017). Nessa etapa é necessário um cuidado para que as bibliotecas reflitam a proporção contida na amostra original (Rani & Sharma, 2017, Hrdlickova, Toloue, & Tian, 2017), e que haja uma preocupação com o número de réplicas a serem sequenciadas (Schurch et al., 2016).

Dessa forma, as informações obtidas durante a etapa de sequenciamento são definidas de acordo com os protocolos utilizados para construção das bibliotecas. Os protocolos mais aplicados são: *single-end* e *paired-end*, que diferem basicamente se será sequenciada uma extremidade dos fragmentos gerados na etapa (iii) ou ambas as extremidades, respectivamente. Os resultados gerados do sequenciador (dados brutos) em formato *.fastq* é utilizado como dados de entrada para início do nosso roteiro (pipeline) de identificação de genes diferenciais. O pipeline é composto pelas etapas: i) análise de qualidade e trimagem dos dados; ii) escolha de uma referência; iii) alinhamento dos fragmentos sequenciados (reads) à referência; iv) quantificação dos transcritos; e v) análise de expressão diferencial. É importante ressaltar que durante a etapa de alinhamento deve-se considerar informações sobre o protocolo utilizado na preparação das bibliotecas.

Dessa forma, esse capítulo tem como objetivo apresentar um pipeline de análise de expressão gênica pareada desenvolvido em linguagem R e aplicado a duas plantas submetidas duas condições de temperatura, 23°C (controle) e 42°C (estresse). O pipeline descrito engloba trimagem dos dados utilizando o software Trimmomatic (Bolger, Lohse, & Usadel, 2014); alinhamento e quantificação dos reads com o Kallisto (Bray, Pimentel, Melsted, & Pachter, 2016); análises de PCA e dendograma a partir da matriz de contagem; análises de expressão diferencial com os pacotes: Sleuth (Pimentel, Bray, Puente, Melsted, & Pachter, 2017) 2017, DESeq2 (Love, Huber, & Anders, 2014) e EdgeR (Robinson, McCarthy, & Smyth, 2010), sendo apresentado de forma detalhada para introduzir os comandos básicos utilizados na análise.

## 2 | METODOLOGIA

### 2.1 Estudo de caso

Como estudo de caso para apresentação de um pipeline de análise de expressão diferencial foi escolhido o sistema modelo *A. thaliana*. Essa espécie foi a primeira planta a ter seu genoma sequenciado no ano de 2000, com cinco cromossomos que compõem aproximadamente 135Mb (Balfagon et al., 2019;)(“TAIR,”). A versão

atual, TAIR10, contém 33.602 genes, incluindo 27.416 genes codantes, 1.359 ncRNA e 4.827 pseudogenes e transposons (“TAIR,”). Além da vantagem de ter um genoma pequeno, bem montado e estudado, a *Arabidopsis thaliana* ainda apresenta uma facilidade em aplicação de técnicas de transformação e tem um ciclo de vida curto (aproximadamente 6 semanas) com rápido crescimento (Agrawal, 2018). Foi estabelecido um experimento de *Arabidopsis thaliana* submetida a duas condições de temperatura: 23°C na situação controle (C) e 42°C em estresse de temperatura (HS). Inicialmente, as plantas foram submetidas a 50  $\mu\text{mol m}^{-2} \text{s}^{-1}$  de iluminação e 23°C de temperatura por 30 dias, seguido de 7h nas condições de controle e estresse citados anteriormente. Posteriormente foi retomada a condição inicial para recuperação da planta (Balfagon, et al., 2019)<sup>1</sup>. A partir desse experimento foram sequenciados tecidos de folhas em triplicata biológica (Balfagon, et al., 2019). Para isso, foi utilizada a plataforma Next-Seq 500 (Illumina) sendo gerados reads single-end (Balfagon, et al., 2019) using 2 Accent 5. Os dados brutos foram baixados no banco de dados NCBI e podem ser encontrados no link <https://www.ncbi.nlm.nih.gov/bioproject/555093>.

O primeiro passo da análise é baixar o arquivo SraRunInfo.tsv. Ele contém todos os links do SRA e as informações das amostras. Para isso, devemos selecionar as amostras de alta temperatura (HS) (SRR9696660, SRR9696664, SRR9696668) e Controle (C) (SRR9696658, SRR9696662, SRR9696666). Posteriormente, clicar em ‘Send to’ -> File -> RunInfo. O arquivo SraRunInfo.tsv é utilizado para realização do download e conversão do arquivo SRA em fastq, através do comando fastqdump, disponível no pacote SRA Toolkit (“SRA Toolkit,”), como mostra o comando abaixo:

```
> setwd(".")
> base_dir <- getwd()
> dados <- read.csv("SraRunInfo.csv", stringsAsFactors=FALSE)
> arquivos <- basename(dados$download_path)
> for(i in 1:length(arquivos)){
  download.file(dados$download_path[i], arquivos[i])
}

> for(a in arquivos) {
  cmd = paste("fastq-dump --split-3", a)
  system(cmd)
}
```

## 2.2 Pré-tratamento dos dados brutos

Os dados brutos foram analisados com o software FastQC (versão 0.11.8) e trimados com o software trimmomatic (versão 0.39). Salientamos que todos os programas externos ao R devem estar previamente instalados e disponibilizados no /usr/bin/. Para o download automático dos programas devem-se seguir os comandos:

```

> cmd = paste("wget
http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.39.zip")
> system(cmd)
> cmd = paste("unzip Trimmomatic-0.39.zip")
> system(cmd)
> cmd = paste("wget https://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.8.zip")
> cmd = paste("unzip fastqc_v0.11.8.zip")
> system(cmd)
> cmd = paste("chmod 755 ",base_dir,"FastQC/fastqc")
> system(cmd)

```

Para análise dos dados e trimagem:

```

>for(a in arquivos){
  cmd = paste(base_dir,"FastQC/fastqc ",a,".fastq ",sep = "")
  system(cmd)
}

>for(a in arquivos){
  cmd = paste("java -jar ",base_dir,"/Trimmomatic-0.39/trimmomatic-0.39.jar SE -threads 10 -trimlog
",a,".trimlog - summary ",a,".summary ",a,".fastq ",a,".trim.fastq ILLUMINACLIP:Trimmomatic-
0.39/adapters/TruSeq2-SE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15
MINLEN:36",sep = "")
  system(cmd)
}

```

## 2.3 Obtenção da matriz de contagem para análise de expressão diferencial

Após a etapa de preparação dos dados foi realizado um alinhamento utilizando o software Kallisto (versão 0.44) utilizando como referência TAIR10. A partir desse alinhamento os transcritos/genes foram quantificados e a matriz de counts foi transferida ao R. Salientamos que o download do software e o índice do genoma devem ser realizados pelos comandos:

```

> cmd = paste("wget https://github.com/pachterlab/kallisto/releases/download/v0.44.0/kallisto_linux-
v0.44.0.tar.gz")
> cmd = paste("tar -xzf kallisto_linux-v0.44.0.tar.gz")
> system(cmd)
> cmd = paste(base_dir,"/kallisto_linux-v0.44.0/kallisto index -i arabidopsis_index
<caminho_transcriptoma.fa>",sep="")
> system(cmd)

```

O comando abaixo reflete a execução da quantificação sobre o transcriptoma da *Arabidopsis thaliana* utilizando o index 'arabidopsis\_thaliana' sobre a lista 'arquivos', que contém os IDs das bibliotecas do SRA.

```

>for(a in arquivos){
  cmd = paste(base_dir,"/kallisto_linux-v0.44.0/kallisto quant -i arabidopsis_index -o ",a,"_kallisto -b
100 -t 10 --single -l 100 -s 0.001 ",a,".trim.fastq",sep="")
  system(cmd)
}

```

Após quantificar todos os arquivos é gerada uma cópia de cada arquivo fora do diretório para ser utilizado na montagem da matriz, como mostra o comando abaixo:

```

>cmd = paste("for file in ls -l -d SRR*;do cp $file/abundance.tsv $file.tsv;done")
>print(cmd)
>system(cmd)

```



Por último é gerada a matriz de contagem dos genes do TAIR10 para ser utilizada na expressão diferencial. Para isso, utilizamos o script `'abundance_estimates_to_matrix.pl'` presente no software Trinity (versão 2.8.6). Os comandos abaixo mostram o download automático da versão mais recente do software, sua descompactação e a geração da matriz, utilizando os arquivos `.tsv` de quantificação gerados pelo kallisto.

```
>cmd = ("wget https://github.com/trinityrnaseq/trinityrnaseq/releases/download/v2.8.6/trinityrnaseq-
v2.8.6.FULL.tar.gz")
>system (cmd)
>cmd = (" tar -xzf trinityrnaseq-v2.8.6.FULL.tar.gz")
>system(cmd)

>lista = paste0(arquivos,".tsv",collapse = " ")
>cmd = paste("perl ",base_dir,"/trinityrnaseq-2.8.6/util/abundance_estimates_to_matrix.pl --
est_method kallisto --gene_trans_map none ",lista,sep=" ")
>print (cmd)

>head(kallisto.isoform.counts.matrix)
```

	SRR9696658.tsv	SRR9696662.tsv	SRR9696666.tsv	SRR9696660.tsv	SRR9696664.tsv	SRR9696668.tsv
AT1G74180.1	67.4293	134.851	1.01137e-05	337.132	202.121	67.4288
AT5G44980.1	15.5	12.5937	7.87109	50.375	23.6133	16.5
AT2G33680.5	129	174.532	1.9115e-06	134.225	1.43061e-07	38
AT4G27300.1	649	1109	678	567	383	296
AT4G03220.1	0	3	3	3	1	6
AT1G74810.3	18.3047	34.9407	0.00015031	0.80261	7.23279	0
AT2G41720.2	285.198	300.788	291.572	446.138	283.682	298.439
AT4G01220.1	250.564	294.753	298	290.35	113.997	147.923
AT5G23220.1	0	0	0	1	0	2

## 2.4 Análise de componentes principais (PCA) e dendograma

A análise de componentes principais (PCA) é um método de redução de dimensionalidade através de transformações ortogonais das variáveis em componentes principais. Nessa análise as variáveis são as amostras, e a variância anotada para o teste é dos transcritos em relação às amostras. O código abaixo consiste na execução do PCA sobre a matriz de contagem. Os comandos utilizados serão descritos na Seção 2.5.2, que realiza a análise de expressão diferencial pelo DESeq2. O último comando `varianceStabilizingTransformation()` gera o objeto `'vsd'` através da normalização da tabela de contagem pelos fatores de dispersão e o tamanho da biblioteca. A variância normalizada de cada variável é gerada e então, é construído o gráfico do PCA (Figura 1 - A) pelo comando `plotPCA()`.

```

>library( "DESeq" )
>countsTable <- read.delim("kallisto.isoform.counts.matrix", header=TRUE, stringsAsFactors=TRUE)
>rownames(countsTable) <- countsTable[,1]
>countsTable <- countsTable[,2:23]
>conds <- factor(c(names(countsTable)))

>countsTable_novo <- apply(countsTable,2,as.integer)
>countsTable_novo[is.na(countsTable_novo)] <- 0

>cds<-newCountDataSet(countsTable_novo,conds)
>cds<-estimateSizeFactors(cds)
>sizeFactors(cds)
>cds <- estimateDispersions(cds,method='blind')
>vsd <- varianceStabilizingTransformation(cds)

>plotPCA(vsd)

```

A análise de dendograma gera um diagrama de árvore que mostra o agrupamento das variáveis de interesse. No exemplo utilizado das amostras o agrupamento é dado por similaridade. Esse algoritmo é derivado de algoritmos de agrupamentos (clusterização).

O código abaixo reflete a Figura 1 - B. Ele consiste na geração do dendograma sobre os dados de contagem através do pacote *ggdendro*, que necessita do pacote *ggplot2*. Utilizando-se da matriz de contagem, encontramos a distancia euclidiana sobre os dados transpostos através do comando *dist()* e, posteriormente, é gerada a árvore através do comando *hclust()*. Por fim, o comando *ggdendrogram()* nos mostra o gráfico rotacionado.

```

> #install.packages("ggdendro")
> #install.packages("ggplot2")
> library("ggplot2")
> library("ggdendro")

> countsTable <- read.delim("kallisto.isoform.counts.matrix", header=TRUE,
stringsAsFactors=TRUE,row.names = 1)
> dd <- dist(t(scale(countsTable)), method = "euclidean")
> hc <- hclust(dd, method = "ward.D2")
> ggdendrogram(hc, rotate = TRUE, theme_dendro = FALSE,
size = 1) + xlab("Amostras") +ylab("Altura")

```

## 2.5 Análise de expressão diferencial utilizando pacotes do R

A partir da matriz de contagem foram realizadas análises de análise de componentes principais (PCA) e dendograma para identificar similaridade entre as amostras. Em seguida foram explorados três pacotes para seleção de genes diferenciais, DESeq2 (versão 1.26.0), Sleuth (versão 0.27.3) e edgeR (versão 3.28.0). Os códigos do DESeq2 e edgeR foram baseados no programa Trinity.

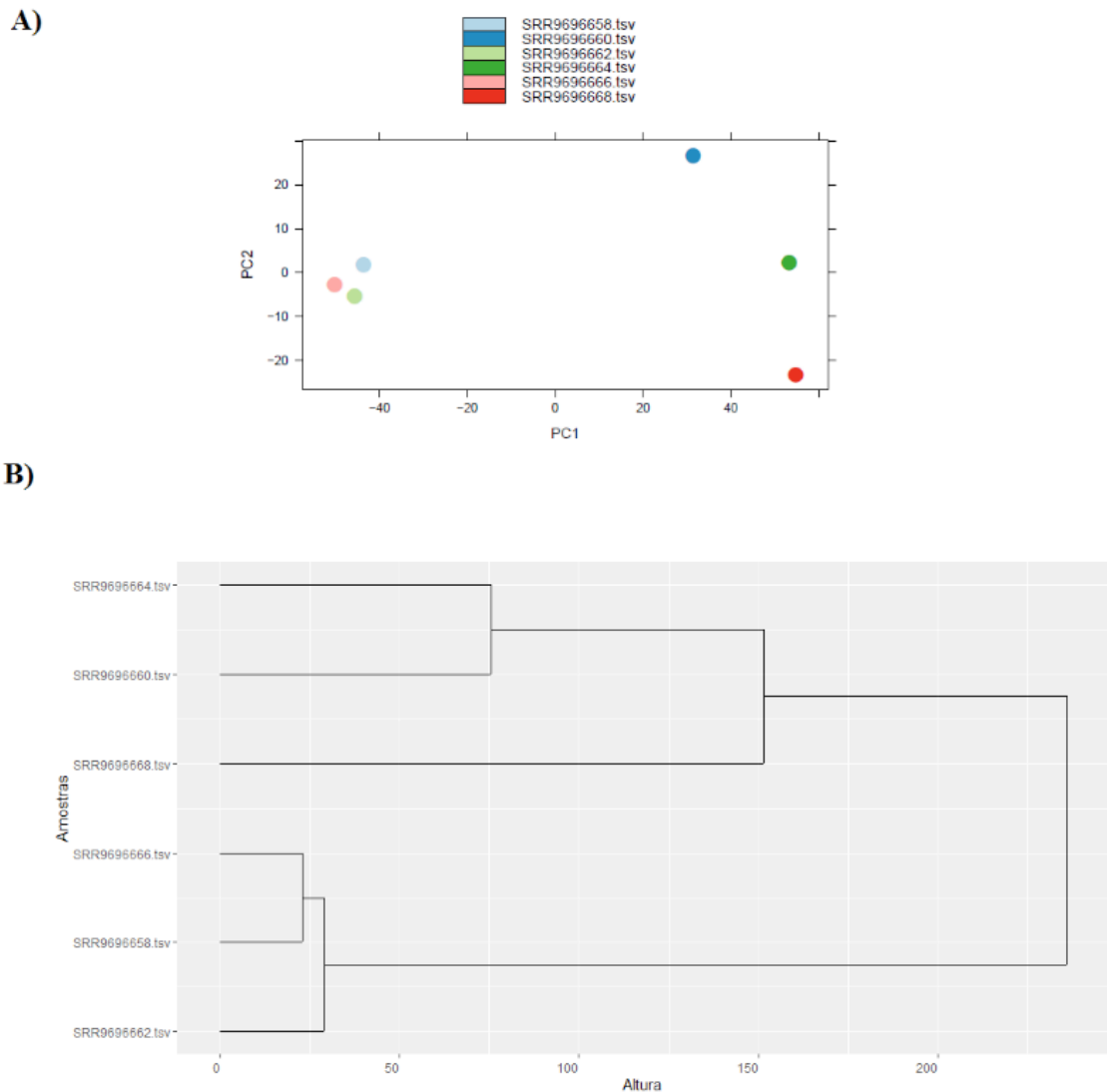


Figura 1. **Resultado PCA e dendograma.** (A) Gráfico da análise de componentes principais (PCA) mostra que as amostras controle (SRR9696658, SRR9696662, SRR9696666) e tratado em alta temperatura (SRR9696660, SRR9696664, SRR9696668) estão agrupados, ou seja, possuem variância semelhante em relação a quantificação dos genes. (B) Dendograma das amostras controle e tratado, mostrando um agrupamento respeitando a condição biológica.

### 2.5.1 EdgeR

Primeiramente é verificado a existência do pacote edgeR, caso não haja ele instalado, ocorre a instalação e a sua chamada.

```
>if (!require(edgeR)) {
  source("https://bioconductor.org/biocLite.R")
  biocLite("edgeR")
  library(edgeR)
}
```

Posteriormente é lida a matriz de quantificação através do comando `'read.table'`. Também ocorre a aproximação inteira dos *counts* através do comando `'round'` e a verificação de no mínimo dois reads counts por linha, através do comando `'rowSums(cpm(rnaseqMatrix) > 1) >= 2'`.

```

>data = read.table("kallisto.isoform.counts.matrix", header=T, row.names=1, com="")
>col_ordering = c(1,2,3,4,5,6)
>rnaseqMatrix = data[,col_ordering]
>rnaseqMatrix = round(rnaseqMatrix)
>rnaseqMatrix = rnaseqMatrix[rowSums(cpm(rnaseqMatrix) > 1) >= 2,]

```

A parte do teste diferencial começa criando a variável *conditions*, que agrupa os dados presentes no arquivo *'kallisto.isoform.counts.matrix'*. A função *DGEList* gera um objeto que contém as contagens inteiras dos genes, um dataframe contendo informações sobre as amostras ou bibliotecas, e um dataframe adicional contendo informações adicionais dos genes. A função *calcNormFactors* normaliza a quantificação do RNA por um conjunto de fatores de escala para os tamanhos da biblioteca que minimizam as alterações do logFC entre as amostras para a maioria dos genes. As dispersões do modelo são calculadas pelo comando *estimateDisp*. Por fim, é executado o teste comparativo pelo comando *exactTest*.

```

>conditions = factor(c(rep("Controle", 3), rep("Tratado", 3)))

>exp_study = DGEList(counts=rnaseqMatrix, group=conditions)
>exp_study = calcNormFactors(exp_study)
>exp_study = estimateDisp(exp_study)
>et = exactTest(exp_study, pair=c("Controle", "Tratado"))
>tTags = topTags(et,n=NULL)

```

A matriz de resultados é extraída e convertido o valor do logFC.

```

>result_table = tTags$table
>result_table = data.frame(sampleA="Controle", sampleB="Tratado", result_table)
>result_table$logFC = -1 * result_table$logFC

```

Os resultados e a matriz são guardados em seus respectivos arquivos através do comando *write.table*.

```

>write.table(result_table, file='edgeR.resultados', sep=' ', quote=F, row.names=T)
>write.table(rnaseqMatrix, file='edgeR.count_matrix', sep=' ', quote=F, row.names=T)

```

Um dos primeiros resultados é a análise do *Volcano plot* (Figura 2), um gráfico que mostra a dispersão dos dados e verifica quais os genes que realmente passaram pelo teste estatístico. O comando abaixo utiliza os dados de logFC (eixo x) e o valor negativo do resultado do teste de expressão ajustado pelo teste de *False Discovery Ratio (FDR)*. Além disso, a adição de cores depende do valor do FDR, sendo que valores inferiores a 0.05 são pintados de vermelho (red), e representam um limite máximo de 5% de falsos positivos no teste, caso contrário de preto (black).

```

>pdf("edgeR.Volcano.pdf")
>plot(result_table$logFC, -1*log10(result_table$FDR), col=ifelse(result_table$FDR<=0.05, "red",
"black"),xlab="logCounts", ylab="logFC", title="Volcano plot", pch=20)
>dev.off()

```

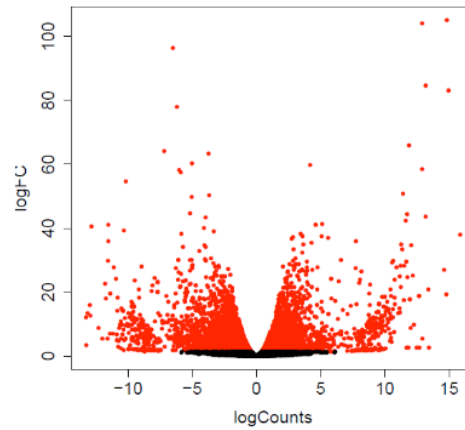


Figura 2. **Volcano plot gerado a partir do resultado do edgeR.** O gráfico mostra a dispersão dos dados e indica através de cores os genes que realmente passaram pelo teste estatístico. Em vermelho estão os genes que passaram pelo teste com valores inferiores a 0.05; e em preto os genes que apresentaram um FDR maior que 0.05.

Os genes significativos do teste são selecionados utilizando o pacote `dplyr()` sobre o campo FDR da tabela `result_table`. Por fim, geramos uma lista do nome dos genes que passaram sobre o teste através de um loop sobre a matriz, gerando a lista `'edgeR_list'`.

```
>edgeR_significant <- dplyr::filter(result_table, FDR <= 0.05)
>edgeR_list <- NULL
>for(i in 1:length(result_table[,1])){
  if(result_table[i,]$FDR <= 0.05){
    edgeR_list[i] <- row.names(result_table)[i]
  }
}
```

```
>head(edgeR_list)
```

```
[1] "AT2G29630.4" "AT2G48060.3" "AT5G12110.1" "AT2G18960.2" "AT2G18960.3" "AT1G12900.4"
```

### 2.5.2 DESeq2

Primeiramente é verificado a existência do pacote DESeq2, caso não haja ele instalado, ocorre a instalação e a sua chamada.

```
>if (! require(DESeq2)) {
  source("https://bioconductor.org/biocLite.R")
  biocLite("DESeq2")
  library(DESeq2)
}
```

Posteriormente são executados os mesmos filtros que a seção 2.5.1.

```

>data = read.table("kallisto.isoform.counts.matrix", header=T, row.names=1, com=")
>col_ordering = c(1,2,3,4,5,6)
>rnaseqMatrix = data[,col_ordering]
>rnaseqMatrix = round(rnaseqMatrix)
>rnaseqMatrix = rnaseqMatrix[rowSums(cpm(rnaseqMatrix) > 1) >= 2,]
>conditions = data.frame(conditions=factor(c(rep("Controle", 3), rep("Tratado", 3))))
>rownames(conditions) = colnames(rnaseqMatrix)

```

A função DESeq realiza a comparação entre as condições, estimando os fatores de escala (*estimateSizeFactors()*), as dispersões (*estimateDispersions()*) e realizando o teste estatístico através de um modelo binomial negativa como hipótese.

```

>ddsFullCountTable <- DESeqDataSetFromMatrix(
  countData = rnaseqMatrix,
  colData = conditions,
  design = ~ conditions)

>dds = DESeq(ddsFullCountTable)
>contrast=c("conditions","Controle","Tratado")
>res = results(dds, contrast)

```

A variável *baseMeanA* e *baseMeanB* guardam o valor médio da expressão de cada gene na condição 'Controle' e 'Tratado', respectivamente. Elas são adicionadas ao dataframe 'res' através do comando *cbind()*. O p-value ajustado dos genes com resultado NA são modificados para 1. Por último é gerado o arquivo de resultados através do comando *write.table()*.

```

>baseMeanA <- rowMeans(counts(dds, normalized=TRUE)[,colData(dds)$conditions == "Controle"])
>baseMeanB <- rowMeans(counts(dds, normalized=TRUE)[,colData(dds)$conditions == "Tratado"])
>res = cbind(baseMeanA, baseMeanB, as.data.frame(res))
>res = cbind(sampleA="Controle", sampleB="Tratado", as.data.frame(res))
>res$padj[is.na(res$padj)] <- 1
>res = as.data.frame(res[order(res$pvalue),])
>write.table(res, file='DESeq2.DE_results', sep=' ', quote=FALSE)

```

A geração do *Volcano plot* (Figura 3) e a lista de genes também são geradas da mesma maneira que na Seção 2.5.1.

```

>pdf("DESeq2_Volcano.pdf")
>plot(res$log2FoldChange, -1*log10(res$padj), col=ifelse(res$padj<=0.05, "red",
"black"),xlab="logCounts", ylab="logFC", title="Volcano plot", pch=20)
>dev.off()

>Deseq2_list <- NULL
>for(i in 1:length(res[,1])){
  if(res[i,]$padj <= 0.05){
    Deseq2_list[i] <- row.names(res)[i]
  }
}

>head(Deseq2_list)

[1] "AT5G12110.1" "AT1G64780.1" "AT1G77510.1" "AT2G29500.1" "AT5G61590.1" "AT1G09750.1"

```

### 2.5.3 Sleuth

Primeiramente é verificada a existência do pacote Sleuth. Caso não haja ele instalado, ocorre a instalação e a sua chamada.



```
if (!require(DESeq2)) {
  devtools::install_github("pachterlab/sleuth")

  library("sleuth")}
```

Primeiramente criamos uma lista das amostras que serão testadas e os caminhos para os diretórios gerados pelo kallisto, já que o sleuth depende dos seus resultados para sua execução.

```
>sample_id <- list('SRR9696658','SRR9696662','SRR9696666',
'SRR9696660','SRR9696664','SRR9696668')

>paths <- list(paste(base_dir,"/SRR9696658",sep=""),
  paste(base_dir,"/SRR9696662",sep=""),
  paste(base_dir,"/SRR9696666",sep=""),
  paste(base_dir,"/SRR9696660",sep=""),
  paste(base_dir,"/SRR9696664",sep=""),
  paste(base_dir,"/sleuth/SRR9696668",sep=""))

>names(paths) <- sample_id
```

Posteriormente é criado um dataframe que depende de um arquivo chamado 'amostras.txt', composto pela descrição das amostras e suas réplicas (mostrado abaixo). Também são selecionados campos específicos das amostras pelo comando 'dplyr::select' e então anexados a variável 'paths' ao dataframe s2c através do comando 'dplyr::mutate'.

```
> am <- read.table(file="amostras.txt",sep="\t",header=TRUE)
> head(am)
```

sample	condition	reps
1	SRR9696658	Controle 1
2	SRR9696662	Controle 1
3	SRR9696666	Controle 1
4	SRR9696660	Tratado 2
5	SRR9696664	Tratado 2
6	SRR9696668	Tratado 2

```
>s2c <- read.table(file.path(base_dir, "amostras.txt"), header = TRUE, stringsAsFactors=FALSE)
>s2c <- dplyr::select(s2c, sample = sample, condition, reps)
>s2c <- dplyr::mutate(s2c, path = paths)
>print(s2c)
>s2c <- data.frame(lapply(s2c, as.character), stringsAsFactors=FALSE)
```

A preparação do teste de expressão diferencial é realizado pelo *sleuth\_prep()*:

```
>#transcrito
>so <- sleuth_prep(s2c, ~condition, extra_bootstrap_summary = TRUE)
>so <- sleuth_fit(so)
>so <- sleuth_wt(so, "conditionTratado")
>models(so)
```

A matriz de resultados é gerada, juntamente com os transcritos significativos, que são guardados na lista 'sleuth\_list'. Os resultados são guardados no arquivo *diferenciais\_sleuth.txt*. O comando *sleuth\_live()* é uma visualização interativa dos dados com o Shiny.

```

> results_table <- sleuth_results(so, test='conditionTratado', test_type = 'wald')
> head(results_table)

  target_id      pval      qval      b      se_b mean_obs var_obs tech_var sigma_sq
smooth_sigma_sq
1 AT5G12110.1 5.137705e-229 1.730584e-224 4.501009 0.1393090 7.058186 6.099318 0.0044212051
0.022568366 0.02468928
2 AT2G29630.4 4.253194e-226 7.163229e-222 -8.871293 0.2763547 3.742500 23.633993 0.0037819679
0.026265685 0.11077587
3 AT1G12900.3 3.413529e-209 2.874533e-205 -8.927552 0.2892355 3.779464 23.956643 0.0192204939
0.038636202 0.10626525
4 AT1G12900.4 3.413529e-209 2.874533e-205 -8.927552 0.2892355 3.779464 23.956643 0.0192204939
0.038636202 0.10626525
5 AT1G20020.2 1.776434e-117 1.196748e-113 -9.461283 0.4106134 4.037494 27.057087 0.0074449443
0.245460135 0.07994148
6 AT5G61590.1 2.968696e-102 1.666626e-98 -2.905480 0.1353273 6.528624 2.546685 0.0031903136
0.014485792 0.02427990
7 AT1G75750.1 5.986875e-83 2.880884e-79 2.735799 0.1417925 7.718472 2.269505 0.0009833116
0.029174350 0.02667726

> sleuth_significant <- dplyr::filter(results_table, qval <= 0.05)
> sleuth_list <- sleuth_significant[,1]
> head(sleuth_list)

[1] "AT5G12110.1" "AT2G29630.4" "AT1G12900.3" "AT1G12900.4" "AT1G20020.2" "AT5G61590.1"

>write.table(sleuth_significant,file="diferenciais_sleuth.txt")
>sleuth_live(so)

>pdf("Sleuth_Volcano.pdf")
>plot(results_table$b, -1*log10(results_table$qval), col=ifelse(results_table$qval<=0.05, "red",
"black"),xlab="log(qval)", ylab="beta", title="Volcano plot", pch=20)
>dev.off()

```

## 2.6 Comparação entre os pacotes para expressão diferencial

Nesta seção serão apresentados os resultados as análises de expressão diferenciais realizadas na seção anterior. A Figura 3 mostra o diagrama de venn dos transcritos diferencialmente expressos entre os três testes diferenciais (sleuth, DESeq2 e edgeR). O digrama foi gerado através do pacote ‘*VennDiagram*’, no qual o código está disponibilizado no script.R no repositório github no link: [https://github.com/Imiguel/DEG\\_Athaliana](https://github.com/Imiguel/DEG_Athaliana).

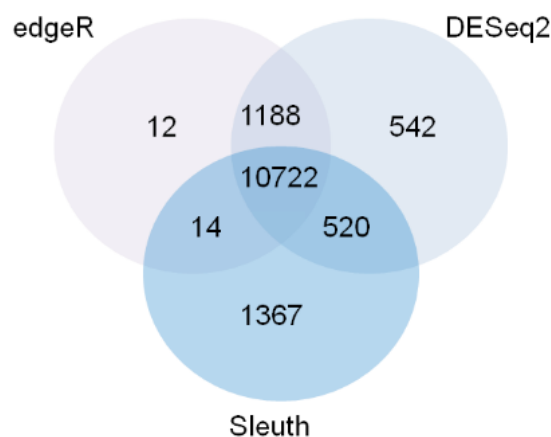


Figura 3. **Comparação entre os genes diferenciais obtidos pelos pacotes utilizados** Na comparação entre os pacotes foram identificados 10722 genes diferenciais em comum (core), sendo que o edgeR identificou 12 genes não detectados nos outros pacotes, DESeq2, 542 e Sleuth 1367 genes diferenciais.

A análise revela um grande core de genes em comum entre os três pacotes analisados (10.722 transcritos). Comparado com o artigo original dos dados, que utilizou o pacote DESeq2 para análise de expressão diferencial, o nosso resultado se aproxima, havendo pequenas diferenças causadas pelo algoritmo de quantificação dos transcritos (*kallisto* versus mapeamento no transcriptoma).

A partir desses resultados, podem ser realizadas análises específicas como enriquecimento do core de genes entre os métodos testados, visto que, teoricamente ele se conserva independente do método estatístico utilizado. Um possível pacote para a análise de enriquecimento de genes, utilizando a linguagem R, é o pacote *topGO()*.

### 3 | ADAPTAÇÃO DO PIPELINE EM DIFERENTES CONTEXTOS

Ao planejar um experimento de RNA-seq para análise de expressão diferencial é importante considerar a quantidade de réplicas que serão sequenciadas, considerando que, em geral, para que todos os genes encontrados sejam relevantes, pede-se um mínimo de 12 réplicas (Schurch, et al., 2016).

Na análise de bioinformática, a primeira tomada de decisão é a escolha de como será realizado o pré-processamento dos dados. No pipeline disponibilizado foi utilizado o software FastQC (Andrews, 2010) para análise de qualidade e o Trimmomatic (Bolger, et al., 2014) para trimagem dos reads. Porém, também pode ser realizada uma análise para descarte de RNA ribossomal com o software SortMeRNA (Carvalho et al., 2019; Kopylova, Noe, & Touzet, 2012).

A segunda etapa importante é escolher o genoma/transcriptoma que possa ser utilizado como referência. Nessa etapa podem ser utilizados os bancos de dados como EnsemblPlants (“EnsemblPlants,”), NCBI (*NCBI Genome*), Phytozome (“Phytozome,”) e alguns específicos para uma espécie como o TAIR (“TAIR,”), desenvolvido para *Arabidopsis thaliana*. Para os casos em que não existe genoma público, é possível realizar a montagem do transcriptoma com os dados sequenciados utilizando, por exemplo, o software Trinity (Haas et al., 2013). Após a escolha da referência é necessário o mapeamento dos reads e a quantificação da expressão, que pode ser realizada a nível de transcrito e/ou de genes. Com esses dados têm-se início a análise de expressão diferencial com os pacotes citados.

É importante considerar que para cada etapa apresentada podem ser utilizados diversos softwares além dos citados (Carvalho, et al., 2019). Dessa forma, o pipeline apresentado pode ser adaptado de acordo com as necessidades do projeto, adequação aos dados e preferência do usuário, sendo de extrema relevância para estudos com plantas.

### 3.1 Disponibilidade de dados e script

O material pode ser baixado em: [doi.org/10.13140/RG.2.2.20790.86086](https://doi.org/10.13140/RG.2.2.20790.86086)

#### AGRADECIMENTOS

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo auxílio financeiro à pesquisa (Processos: 150977/2019-0; 140869/2016-6) e também ao Centro de Computação em Engenharia e Ciências (Processo FAPESP/Cepid: 2013/08293-7).

#### REFERÊNCIAS

- Agrawal, S. (2018). *Arabidopsis thaliana as a Model Organism to Study Plant-Pathogen Interactions*. Springer.
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data Retrieved 04/02/2019, from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Balfagon, D., Sengupta, S., Gomez-Cadenas, A., Fritschi, F. B., Azad, R. K., Mittler, R., & Zandalinas, S. I. (2019). Jasmonic Acid Is Required for Plant Acclimation to a Combination of High Light and Heat Stress. *Plant Physiol*, 181(4), 1668-1682. doi: 10.1104/pp.19.00956
- Bazzo, B. R., de Carvalho, L. M., Carazzolle, M. F., Pereira, G. A. G., & Colombo, C. A. (2018). Development of novel EST-SSR markers in the macauba palm (*Acrocomia aculeata*) using transcriptome sequencing and cross-species transferability in Arecaceae species. *BMC Plant Biol*, 18(1), 276. doi: 10.1186/s12870-018-1509-9
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 34(5), 525-527. doi: 10.1038/nbt.3519
- Carvalho, L. M. d., Borelli, G., Camargo, A. P., Assis, M. A. d., Ferraz, S. M. F. d., Fiamenghi, M. B., Carazzolle, M. F. (2019). Bioinformatics applied to biotechnology: A review towards bioenergy research. *Biomass and Bioenergy*, 123, 195-224. doi: 10.1016/j.biombioe.2019.02.016
- Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., Wilkins, T. A., . . . Burow, M. D. (2015). Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut, *Arachis hypogaea* L. *Mol Genet Genomics*, 290(3), 1169-1180. doi: 10.1007/s00438-014-0976-4
- Edwards, D., & Batley, J. (2004). Plant bioinformatics: from genome to phenome. *Trends Biotechnol*, 22(5), 232-237. doi: 10.1016/j.tibtech.2004.03.002 S0167-7799(04)00076-9 [pii] EnsemblPlants. from <https://plants.ensembl.org/index.html>
- Haas, B., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P., Bowden, J., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 8(8). doi: 10.1038/nprot.2013.084
- Hrdlickova, R., Toloue, M., & Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA*, 8(1). doi: 10.1002/wrna.1364

- Kopylova, E., Noe, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24), 3211-3217. doi: 10.1093/bioinformatics/bts611
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550. doi: 10.1186/s13059-014-0550-8 s13059-014-0550-8 [pii]
- Nascimento, L. C., Yanagui, K., Jose, J., Camargo, E. L. O., Grassi, M. C. B., Cunha, C. P., . . . Carazzolle, M. F. (2019). Unraveling the complex genome of *Saccharum spontaneum* using Polyploid Gene Assembler. *DNA Res*, 26(3), 205-216. doi: 10.1093/dnares/dsz001
- NCBI Genome. Retrieved from <https://www.ncbi.nlm.nih.gov/genome/>
- Pimentel, H., Bray, N. L., Puente, S., Melsted, P., & Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods*, 14(7), 687-690. doi: 10.1038/nmeth.4324
- Rani, B., & Sharma, V. K. (2017). Transcriptome profiling: methods and applications- A review. *Agricultural Reviews*, 38(4), 271-281. doi: 10.18805/ag.R-1549
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. doi: 10.1093/bioinformatics/btp616
- Schurch, N. J., Schofield, P., Gierlinsk, M., Cole, C., Sherstnev, A., Singh, V., Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22, 839–851. doi: 10.1261/rna.053959.115
- SRA Toolkit. from <https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/>
- Phytozome. from <https://phytozome.jgi.doe.gov/pz/portal.html>
- TAIR. from <https://www.arabidopsis.org/>

## MODELAGEM DE CRESCIMENTO DE CANA-DE-AÇÚCAR E CANA ENERGIA SOB O ESTÍMULO DE REGULADOR DE CRESCIMENTO

Data de aceite: 12/02/2020

### Luís Guilherme Furlan de Abreu

Universidade Estadual de Campinas – Instituto de Biologia - Departamento de Genética, Evolução e Bioagentes – Laboratório de Genômica e Bioenergia - LGE, Campinas, São Paulo, Brasil

### Lucas Miguel de Carvalho

Universidade Estadual de Campinas – Instituto de Biologia - Departamento de Genética, Evolução e Bioagentes – Laboratório de Genômica e Bioenergia - LGE, Campinas, São Paulo, Brasil

### Maria Carolina de Barros Grassi

Universidade Estadual de Campinas – Instituto de Biologia - Departamento de Genética, Evolução e Bioagentes – Laboratório de Genômica e Bioenergia - LGE, Campinas, São Paulo, Brasil

### Gonçalo Amarante Guimarães Pereira

Universidade Estadual de Campinas – Instituto de Biologia - Departamento de Genética, Evolução e Bioagentes – Laboratório de Genômica e Bioenergia - LGE, Campinas, São Paulo, Brasil

**RESUMO:** Um dos maiores desafios a serem superados é garantir o suprimento sustentável de água, energia e alimento a toda a população mundial, reduzindo impactos negativos no ambiente. Os avanços recentes na biotecnologia vêm revolucionando o setor sucroenergético através do desenvolvimento de novas variedades de cana para geração

de energia, e que apresenta grande potencial para garantir alta produtividade de biomassa, consequentemente açúcar e etanol. Entretanto, uma forma de melhorar a produtividade da cana é através do uso de reguladores químicos de crescimento; por exemplo, se destaca o Ethephon. O Ethephon é um versátil regulador, que atua na via do hormônio vegetal etileno, promovendo o crescimento das plantas. Neste contexto, a avaliação e monitoramento do crescimento das plantas podem ser factíveis através de modelos matemáticos capazes de prever o padrão de crescimento destas plantas ao longo do tempo. Este capítulo teve como objetivo testar um modelo matemático capaz de prever o crescimento de variedades de cana-de-açúcar e cana energia, tratadas com Ethephon. Assim, foi utilizado software R, para o desenvolvimento de comandos necessários para a criação de um modelo de crescimento não-linear. Além disso, mostramos que o modelo matemático a ser usado, pode prever o crescimento da altura da planta em qualquer tempo desejado (valor  $t$ ). Mostramos um modelo estatístico a ser seguido, embora nossos resultados não mostrem diferença significativa entre o Ethephon e controle (sem Ethephon) nas duas variedades utilizadas. Verificamos uma diferença no padrão de crescimento entre cana-de-açúcar e cana energia.

**PALAVRAS-CHAVE:** Cana-de-açúcar; Cana



energia; Crescimento; Ethephon; Modelagem matemática.

**ABSTRACT:** One of the biggest challenges to overcome is ensuring the sustainable supply of water, energy and food to the entire world population, reducing negative impacts on the environment. Recent advances in biotechnology have been revolutionizing the sugarcane industry through the development of new sugarcane varieties for power generation, and which has great potential to ensure high productivity of biomass, hence sugar and ethanol. However, one way to improve sugarcane productivity is through the use of chemical growth regulators; for example, the Ethephon stands out. Ethephon is a versatile regulator that acts in the pathway of the plant ethylene hormone, promoting plant growth. In this context, the evaluation and monitoring of plant growth may be feasible through mathematical models capable of predicting the growth pattern of these plants over time. This chapter aimed to test a mathematical model capable of predicting the growth of sugarcane and energy cane varieties treated with Ethephon. Thus, R software was used to develop the necessary commands to create a nonlinear growth model. In addition, we show that the mathematical model to be used can predict growth of plant height at any desired time ( $t$  value). We show a statistical model to follow, although our results show no significant difference between Ethephon and control (without Ethephon) in the two varieties used. We found a difference in the growth pattern between sugar cane and energy cane.

**KEYWORDS:** Sugarcane; Energy cane; Growth; Ethephon; Mathematical modeling.

## 1 | INTRODUÇÃO

Uma das maiores barreiras a serem superadas para os países em desenvolvimento é assegurar o suprimento de recursos naturais essenciais – água, energia e alimento – sem prejudicar a qualidade de vida, conservação dos ecossistemas e desenvolvimento econômico para atuais e futuras gerações (Ringler et al., 2013). Sem mudança de hábitos e adoção de novas tecnologias, haverá significativo aumento na demanda por água (55%), energia (80%) e alimentos (60%), o que poderá agravar o cenário atual (Ringler et al., 2013). O consumo de energia baseada em combustíveis fósseis (carvão e petróleo) no mundo é alarmante e aumentou 52% entre 1990 e 2015 (Ritchie & Roser, 2017), principalmente no setor de serviços e transporte, agravando o cenário de aquecimento global com o aumento da emissão de gases do efeito estufa (principalmente CO<sub>2</sub>).

Em paralelo, recentes avanços nas tecnologias industriais e na biotecnologia têm possibilitado a produção e a fermentação de açúcares solúveis (transformação de energia) gerados a partir da desconstrução da biomassa (celulose e hemicelulose), produzindo o chamado etanol de segunda geração (Matsuoka, 2017; Santos et al., 2016). Para o Brasil esta é uma estratégia vantajosa, pois a cana-de-açúcar (*Saccharum spp.*) é capaz de fornecer biomassa de baixo custo, sendo aproveitado

para a produção de etanol com potencial de aumentar a produção nacional em 40%. A cana-de-açúcar é considerada uma das matérias-primas mais apropriadas para o setor de biocombustíveis devido ao grande potencial de acumulação de biomassa e sacarose. O desenvolvimento de tecnologias para a produção de biocombustíveis, a busca por cultivares capazes de produzir quantidades maciças de biomassa lignocelulósica, mantendo alto o rendimento de açúcar, tem revolucionado o melhoramento da cultura de cana. A seleção de híbridos com alta quantidade de biomassa resultou em indivíduos com 180 toneladas de peso úmido por hectare (ha), duas vezes mais que a cana-de-açúcar convencional (92 toneladas de peso úmido por ha) (Carvalho-Netto et al., 2014). Esses híbridos são conhecidos como cana energia. Mesmo com o equilíbrio de carbono orientado mais para a formação da parede celular (16-30% de teor de fibra) e não tanto ao acúmulo de açúcares solúveis (12% de açúcar), a cana energia consegue superar a cana convencional na produção de açúcares dada a enorme produção de biomassa (Santos et al., 2016).

Por ser propagada vegetativamente, o conhecimento do processo de brotação e crescimento inicial da cana, bem como nos fenômenos que interferem, reveste-se de grande importância para a cultura, pelo fato de que um canavial será explorado por um período médio de cinco anos (Casagrande, 1991). O crescimento de um canavial consiste na medida sequencial do acúmulo da matéria orgânica na planta e condições morfológicas da planta em diferentes intervalos de tempo (Magalhães, 1985). O uso de reguladores químicos pode auxiliar no crescimento das plantas. Os reguladores de crescimento podem ser classificados em cinco grupos, auxinas, giberelinas, citocininas, etileno e ácido abscísico. Um dos modos de utilização do etileno é através da aplicação do composto químico comercial Ethephon. O Ethephon é um versátil regulador de crescimento, sendo rapidamente metabolizado por plantas através da hidrólise, produzindo gás etileno, ácido fosfórico e ácido clorídrico; produtos não tóxicos e encontrados naturalmente em todas as plantas (Solomon et al., 1988).

Neste contexto, com o intuito de evitar perdas na produtividade, modelagens matemáticas capazes de prever o padrão de crescimento das plantas são fundamentais para culturas com alto valor econômico, como no caso a cana-de-açúcar e cana energia. Nos últimos anos, com os avanços dos softwares estatísticos e o ajuste do modelo não-linear de crescimento, uma ampla variedade de modelos matemáticos podem ser exploradas (Pinheiro & Bates 2000; Ritz & Streibig 2008). O objetivo deste capítulo é demonstrar a metodologia de análise de um modelo matemático de crescimento, assim como análises estatísticas utilizando software R. Utilizamos os dados iniciais de crescimento da altura ao longo do tempo de cana-de-açúcar e cana energia, tratadas com Ethephon e apenas água como controle.

## 2 | MATERIAL E MÉTODOS

### 2.1 Material vegetal e condições do experimento

O material vegetal utilizado para o experimento foi obtido no Centro de Tecnologia Canavieira (CTC) e na Escola Superior de Agricultura Luiz de Queiroz. Cultivar comercial de cana-de-açúcar (CTC4) e cana energia (Vertix 12) com aproximadamente 10 meses de idade, foram utilizadas neste estudo. O experimento foi realizado em casa de vegetação da Universidade Estadual de Campinas (UNICAMP), Campinas, Brasil (22°49'09.7"S, 47°04'16.4"W) em novembro de 2018. Foram utilizados os colmos +3 ao +14 das duas variedades, os colmos foram padronizados em comprimento, diâmetro e peso para garantir a mesma quantidade de reserva nutritiva. Além disso, os colmos obtidos para os experimentos foram provenientes de plantas individuais, de diferentes perfilhos. Estes colmos foram aleatoriamente imersos em solução de Ethephon (100 ppm) e água como controle, por *overnight*. Em seguida, foram distribuídos em bandejas de plásticos (38 cm de comprimento × 29 cm de largura × 6,5 cm de altura) contendo 2,5 litros de vermiculita expandida super fina (Global Minérios) e após ao plantio, foram cobertos com 1,4 litros de vermiculita, assegurando, desta forma, que todas as bandejas apresentasse a mesma quantidade de vermiculita. Os experimentos foram conduzidos em blocos casualizados em parcelas subdivididas. A irrigação ocorreu diariamente para que não houvesse estresse hídrico. A brotação e, conseqüentemente, o crescimento das plântulas foram analisados diariamente partir da medida do comprimento da parte aérea, diariamente.

### 2.2 Organização da planilha para análise no R

A planilha de dados consiste em 20 observações para cada uma das 18 plantas (réplicas técnicas), tanto para cana-de-açúcar (CTC4) quanto para cana energia (Vertix 12), com ambas em dois momentos: com Ethephon e o controle. Um total de 1440 anotações foi gerado.

Para a modelagem de crescimento foi anotado a média dos pontos dos dias das plantas que foram tratadas com Ethephon e ou controle para cada uma das variedades. Ao final temos quatro condições (CTC4\_controle, CTC4\_ethephon, VTX12\_controle e VTX12\_ethephon). Os dados estão resumidos nos vetores abaixo, que serão utilizados em toda a análise.

```

> setwd(".")
> vtx12_eth <- c(0,0,0,0.366666667,0.85,1.672222222, 2.616666667, 6.716666667,
7.605555556, 9.288888889, 10.71666667, 13.71666667, 14.72777778,
15.86111111, 16.97777778, 19.37777778, 21.14444444, 25.16666667,
25.83333333)

> vtx12_ct <- c(0,0.027777778, 0.061111111, 0.622222222, 1.044444444,
2.105555556, 3.533333333, 8.083333333, 8.905555556, 10.63333333,
12.62222222, 16.72222222, 18.17222222, 19.15, 20.78888889, 24.74444444,
27.25555556, 32.35, 34.12777778)

> ctc4_ct <- c(0,0,0,0.05,0.13125,0.25625,0.36875,0.493333333,0.625,
0.7875,1.46875,1.76875, 2.04375, 2.33125, 3.24375,3.9125,6.2125, 7.4375)

> ctc4_eth <- c(0,0,0, 0.129411765, 0.217647059, 0.229411765, 0.417647059,
0.8, 0.894117647,1.3, 1.617647059,3.476470588, 3.452941176, 4.252941176,
5.317647059,7.847058824, 9.464705882,14.09411765, 16.45294118)

> dias <- c(1,2,3,4,5,6,7,8,12,13,14,15,18,20,22,24,26,28,30)

```

### 2.3 Pacotes estatísticos e scripts

Os pacotes que foram utilizados para a modelagem e análise dos dados estão resumidos abaixo.

1) *nls()*: Determina as estimativas de mínimos quadrados não lineares dos parâmetros de um modelo não linear. Essa função espera como parâmetro uma fórmula de modelagem não linear, utilizaremos a função de regressão logística  $P(t) = \frac{M}{1+e^{P_0+k*t}}$ ; um dataset de entrada e, dados iniciais para as interações do modelo, que no nosso caso seriam os valores de  $M$ ,  $P_0$  e  $k$ , já que  $t$  seria o valor dos dias observados (vetor 'dias').

2) *shapiro.test()*: utilizado para verificar o tipo de distribuição de probabilidade dos dados.. Sua hipótese nula  $H_0$  é de que os dados não apresentam uma distribuição normal, logo, para que o dado seja normal, devemos verificar se o p-value do teste é superior a o valor de corte (utilizamos 0.05).

3) *wilcox.test()*: função de comparação estatística para dados não normais.

4) *t.test()*: função de comparação estatística para dados normais, realizado através do test-t de Student. Será utilizado em dados não pareados, já que o desenho foi desenhado dessa maneira.

## 3 I APLICAÇÃO DA MODELAGEM DE CRESCIMENTO

O primeiro passo é observar que o crescimento vegetativo tanto da cana-de-açúcar como a cana energia é não linear e, para isso, geramos o gráfico de dispersão (*scatter plot*) do crescimento médio da planta (tratado com Ethephon e controle) em função dos dias (Figura 1). O código para configurar a figura consiste em utilizar a função *par()* para gerar uma matriz de malhas de figura, que no nosso caso, foi uma 2x2. Além disso, a função *plot()* é utilizada para gerar os gráficos.

```

> setwd(".")
> windows()
> setwd("../Ethephon/")
> par(mfrow=c(2,2))
> plot(dias,vtx12_ct,las=1,pch=16,main="VTX12 - Controle")
> plot(dias,vtx12_eth,las=1,pch=16,main="VTX12 - Ethephon")
> plot(dias,ctc4_ct,las=1,pch=16,main="CTC4 - Controle")
> plot(dias,ctc4_eth,las=1,pch=16,main="CTC4 - Ethephon")

```

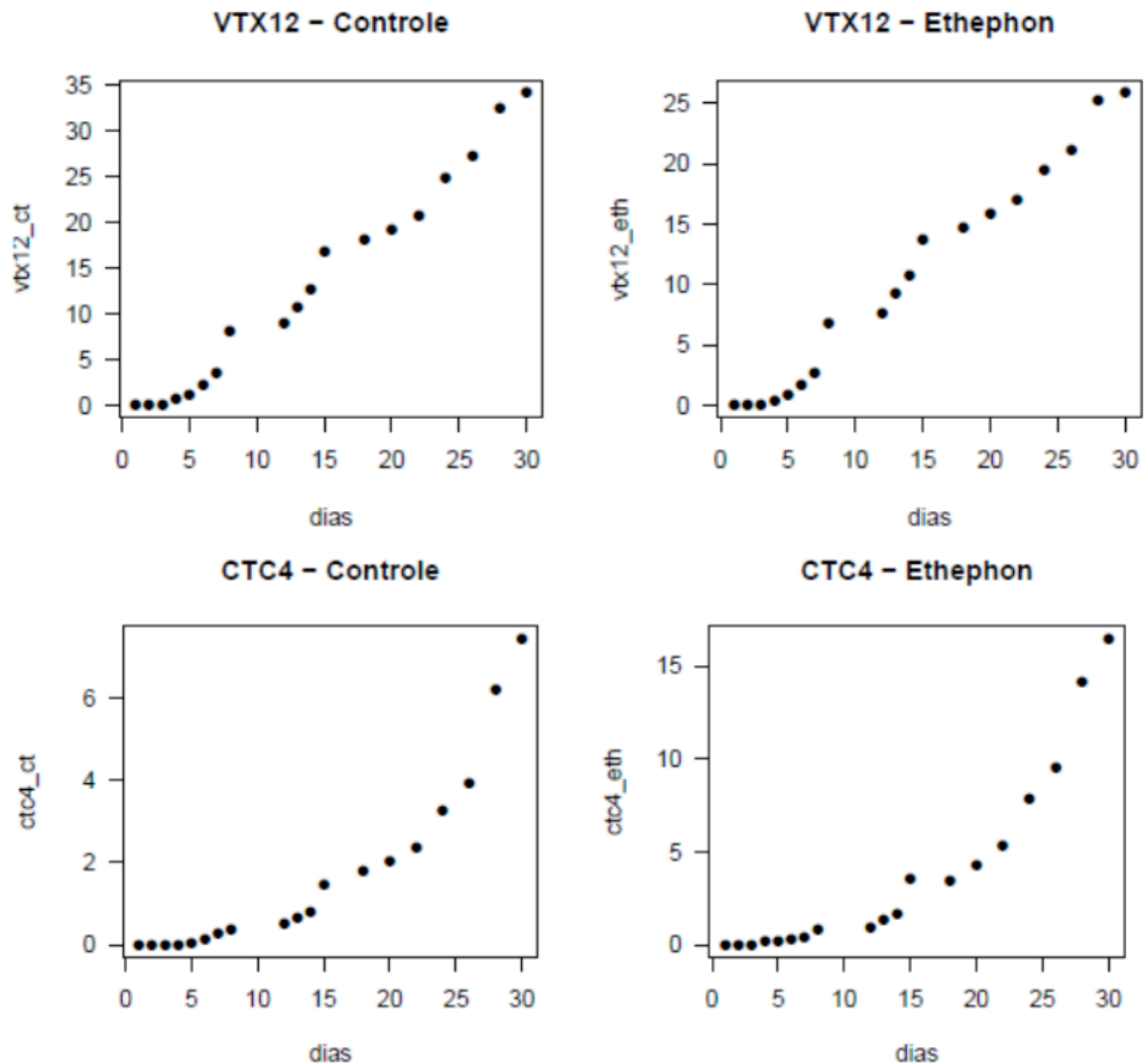


Figura 1. Gráfico de dispersão do crescimento de cada tratamento em função do tempo, em dias. Com esse tipo de gráfico fica evidente que o modelo de crescimento é não linear.

A execução da equação de modelagem se baseia em linearizar os pontos através de uma equação sigmod. Os comandos que geram os modelos do crescimento vegetativo das plantas tratadas com Ethephon e ou controle. O comando `summary()` consegue enunciar os dados e valores encontrados pelo modelo, já o comando `coef()` anota os coeficientes gerados no modelo. É apresentado abaixo o código para a geração do modelo de cana energia com Ethephon (vetor 'vtx12\_eth').

```

> logisticModel1 <- nls(vtx12_eth~M/(1+exp(Po+k*dias)), start=list(Po=0, k=-0.21, M=17))
> summary(logisticModel)
Formula: vtx12_eth ~ M/(1 + exp(Po + k * dias))

Parameters:
  Estimate Std. Error t value Pr(>|t|)
Po 3.23501  0.32066  10.089 2.43e-08 ***
k -0.19302  0.02788  -6.923 3.43e-06 ***
M 26.11089  2.09877  12.441 1.22e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.704 on 16 degrees of freedom

Number of iterations to convergence: 13
Achieved convergence tolerance: 4.901e-06

> coef(logisticModel1)
      Po      k      M
3.2350056 -0.1930236 26.1108931

```

Como resultados observados no modelo, notamos que os valores de  $M$ ,  $P_0$  e  $k$  foram estatisticamente significativos (ambos com  $p\text{-value} \leq 0.05$ ). Ainda nota-se que, após 13 iterações, o modelo atingiu a consistência estatística ( $p\text{-value} = 4.901e-06 \leq 0.05$ ). Com os valores gerados pelo comando `coef()`, chegamos a fórmula da modelagem para a cana energia:

$$P(t) = \frac{26.1109}{1 + e^{3.235 - 0.1930 * t}}$$

Com a fórmula da modelagem de crescimento, podemos prever qual será o tamanho da planta após  $t$  dias. A escolha do modelo é importante para que a predição possa ser analisada, já que alguns modelos tomam como hipótese um crescimento contínuo, como o modelo linear, e outros um plateau após  $t$  dias, como o sigmod. Os modelos para as outras condições de interesse poderão ser gerados através das linhas de comando abaixo.

```

> logisticModel2 <- nls(vtx12_ct~M/(1+exp(Po+k*dias)), start=list(Po=0, k=-0.21, M=17))
> logisticModel3 <- nls(ctc4_ct~M/(1+exp(Po+k*dias)),
  start=list(Po=5, k=-0.5, M=10),
  control = list(maxiter = 100,
  minFactor = 1/4096),trace = TRUE)
> logisticModel4 <- nls(ctc4_100~M/(1+exp(Po+k*dias)),
  start=list(Po=5, k=-0.5, M=30),
  control = list(maxiter = 50,minFactor = 1/4096),
  trace=TRUE)

```

A partir de cada modelagem foi gerado o gráfico de dispersão, juntamente com a curva de predição. A Figura 2 resume a curva de modelagem diante as quatro comparações.

```

> windows()
> par(mfrow=c(2,2))
> x1 <- seq(min(dias), max(dias), length=100)

```



```

> y1 <- predict(logisticModel1, list(dias=x1))
> plot(dias,vtx12_ct,las=1,pch=16, main="VTX12 -Ethephon")
> points(x1, y1, type='l', col='blue')

> x2 <- seq(min(dias), max(dias), length=100)
> y2 <- predict(logisticModel2, list(dias=x2))
> plot(dias,vtx12_eth,las=1,pch=16, main="VTX12 - Controle")
> points(x2, y2, type='l', col='blue')

> x3 <- seq(min(dias), max(dias), length=100)
> y3 <- predict(logisticModel3, list(dias=x3))
> plot(dias,ctc4_ct,las=1,pch=16, main="CTC4 - Controle")
> points(x3, y3, type='l', col='blue')

> x4 <- seq(min(dias), max(dias), length=100)
> y4 <- predict(logisticModel4, list(dias=x4))
> plot(dias,ctc4_eth,las=1,pch=16, main="CTC4 - Ethephon")
> points(x4, y4, type='l', col='blue')

```

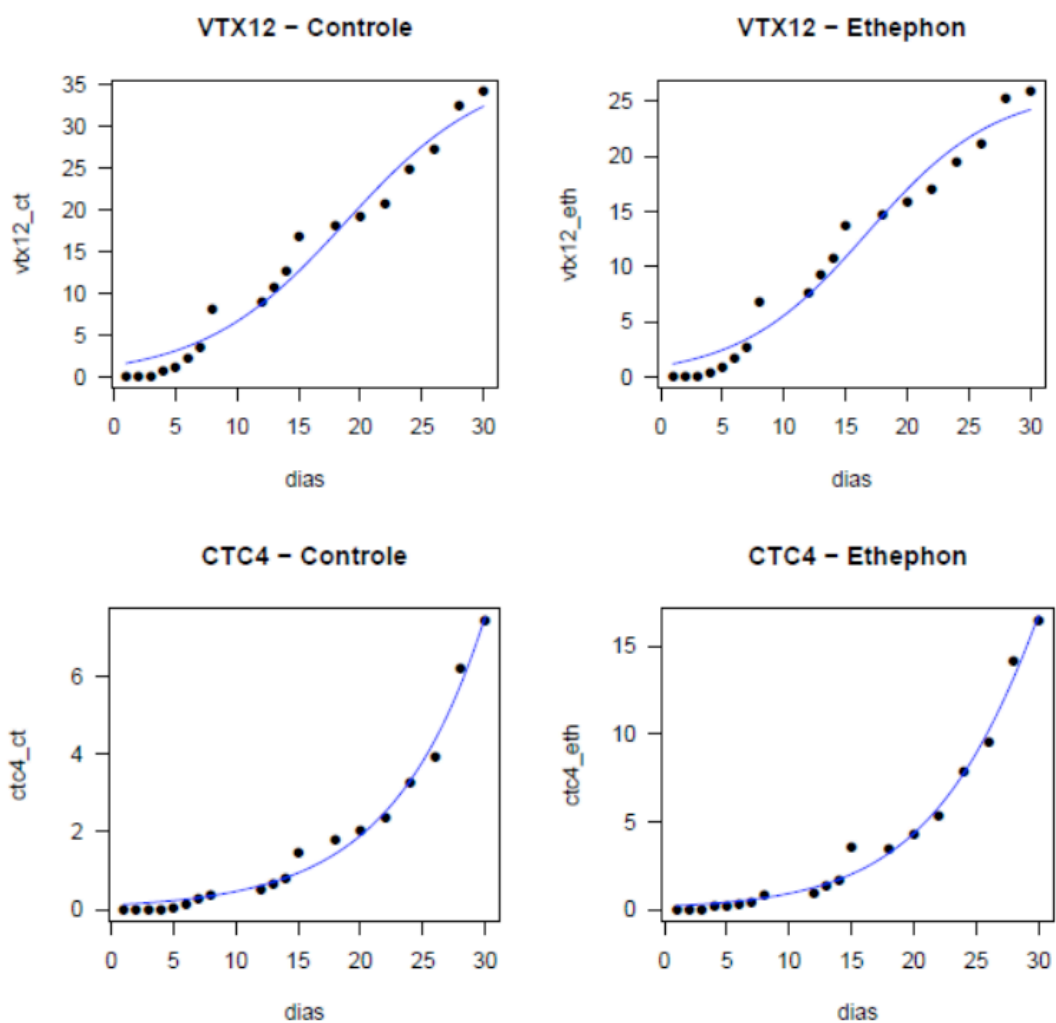


Figura 2. Modelagem realizada através de curvas sigmoids, representada por uma linha contínua azul, em relação ao gráfico de dispersão dos dados originais. Em todos os modelos houve significância estatística para o teste ( $p\text{-value} \leq 0.05$ ).

## 4 | COMPARAÇÃO ESTATÍSTICA

A análise estatística comparativa foi executada para dizer se a aplicação do Ethephon teve de fato significância estatística, tanto na comparação interna a

variedade (VTX12 e CTC4) quanto entre as condições (VTX12 versus CTC4) para ambas as condições de aplicação do Ethephon. Primeiramente vamos realizar o teste de Shapiro para verificar se há normalidade nos dados. Os códigos abaixo resumem a aplicação do teste.

```
> shapiro.test(ctc4_ct)
      Shapiro-Wilk normality test

data:  ctc4_ct
W = 0.76878, p-value = 0.0004163
> shapiro.test(ctc4_eth)
      Shapiro-Wilk normality test

data:  ctc4_eth
W = 0.7621, p-value = 0.0003383
> shapiro.test(vtx12_ct)
      Shapiro-Wilk normality test

data:  vtx12_ct
W = 0.90627, p-value = 0.06324
> shapiro.test(vtx12_eth)
      Shapiro-Wilk normality test

data:  vtx12_eth
W = 0.90578, p-value = 0.06196
```

Com a análise dos resultados notamos que os dados oriundos de cana-de-açúcar não apresentam normalidade ( $p\text{-value} < 0.05$ ), diferente da variedade de cana energia, que apresenta normalidade ( $p\text{-value} \geq 0.05$ ). Com este resultado, iremos aplicar a função `t.test()` para a comparação dados de cana energia tratada com Ethephon e controle. Contudo, para os dados de cana-de-açúcar iremos utilizar a função `wilcox.test()` para compará-las. A comparação feita entre os dados de cana-de-açúcar versus cana energia foi realizada através da função `wilcox.test()`.

```

> res <- wilcox.test(ctc4_ct,ctc4_eth,exact = FALSE)
> res
Wilcoxon rank sum test with continuity correction

data: ctc4_ct and ctc4_eth
W = 143, p-value = 0.2786
alternative hypothesis: true location shift is not equal to 0

> res <- t.test(vtx12_ct,vtx12_eth,exact = FALSE,paired = FALSE)
> res
Welch Two Sample t-test

data: vtx12_ct and vtx12_eth
t = 0.76436, df = 33.969, p-value = 0.4499
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-4.217926  9.303306
sample estimates:
mean of x mean of y
12.68158  10.13889

> res <- wilcox.test(ctc4_ct,vtx12_ct,exact = FALSE)
> res
Wilcoxon rank sum test with continuity correction

data: ctc4_ct and vtx12_ct
W = 72, p-value = 0.001597
alternative hypothesis: true location shift is not equal to 0

> res <- wilcox.test(ctc4_eth,vtx12_eth,exact = FALSE)
> res
Wilcoxon rank sum test with continuity correction

data: ctc4_eth and vtx12_eth
W = 109.5, p-value = 0.03919
alternative hypothesis: true location shift is not equal to 0

```

A análise dos resultados mostra que a comparação entre os dados de cana-de-açúcar controle (ctc4\_ct) versus Ethephon (ctc4\_eth) não apresentou significância estatística (p-value = 0.2786  $\geq$  0.05). Mesmo resultado apresentado pela variedade de cana energia, onde o tratamento com Ethephon (vtx12\_eth) versus controle (vtx12\_ct) não apresentou significância estatística (p-value = 0.499  $\geq$  0.05).

Os resultados comparativos entre as variedades, tanto no controle como no tratamento com Ethephon, deram significância estatística. Ao compararmos as duas variedades tanto no controle (ctc4\_ct versus vtx12\_ct) como no tratamento com Ethephon (ctc4\_eth versus vtx12\_eth) nota-se que há significância estatística, com p-value igual a 0.001597 e 0.03919, respectivamente. Com este resultado podemos concluir que há uma diferença significativa no crescimento entre as variedades, havendo superioridade na linhagem VTX12 de cana energia.

Por último, mostramos a dispersão dos dados através de um gráfico de boxplot (Figura 3). A partir desta figura podemos notar a diferença do crescimento entre as variedades de cana.

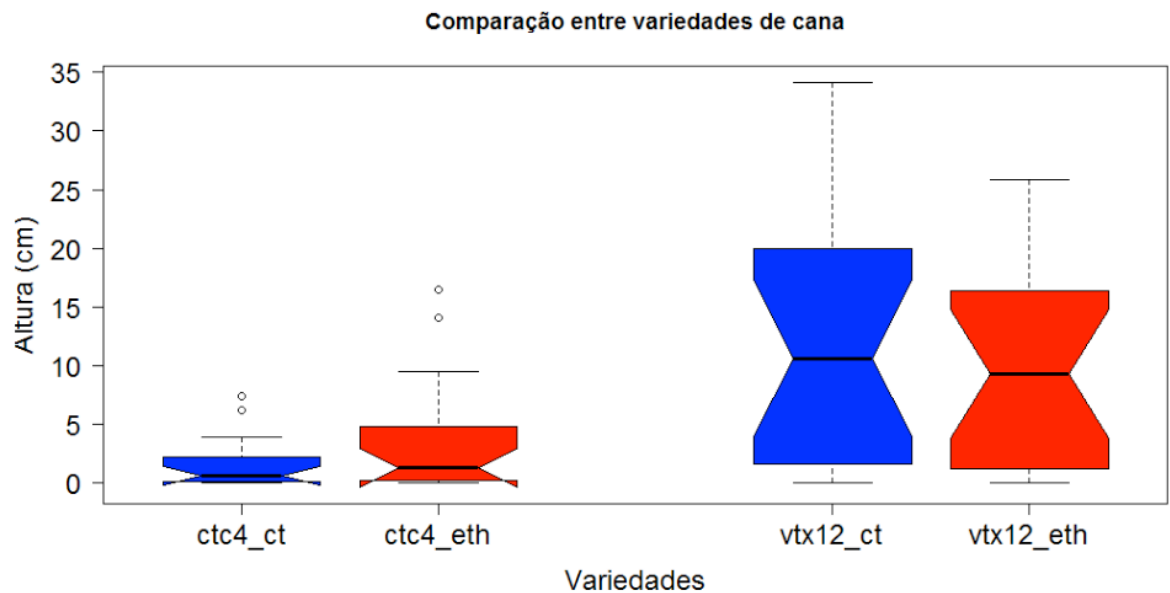


Figura 3. Boxplot comparativo entre as variedades e seus tratamentos. Significado dos tratamentos - ctc4\_ct: cana-de-açúcar controle; ctc4\_eth: cana-de-açúcar com Ethephon; vtx12\_ct: cana energia controle; vtx12\_eth: cana energia com Ethephon.

## ACESSIBILIDADE AOS DADOS E SCRIPT

O *script* e a planilha contendo os dados brutos do experimento podem ser acessados pelo link: <https://github.com/lmigueel/Modelo-Ethephon>.

## AGRADECIMENTOS

Este estudo foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo número: 2018.10315-2.

## REFERÊNCIAS

Carvalho-Netto O., Bressiani J. A., Soriano H L., Fiori C S., Santos J M., Barbosa G VS., Xavier M A., Landell M GA., Pereira G AG. The potencial of the energy cane as the mais biomass crop for the cellulosic industry. *Chemical and Biological Technologies in Agriculture* 2014, 1:20.

Casagrande, A.A. Tópicos de morfologia e fisiologia da cana-de-açúcar. Jaboticabal: FUNEP, 1991. 157p.

Magalhães, A. C. N. Análise quantitativa do crescimento. In: FERRI, M. G. *Fisiologia Vegetal*. 2. ed. São Paulo: Editoras EPU; EDUSP, 1985. v. 1, p. 331- 350.

Matsuoka, S. (2017). Free fiber level drives resilience and hybrid vigor in energy cane. *Journal of Scientific Achievements* 2: 1-35.

Pinheiro, J. & Bates, D. (2000) *Mixed-Effects Models in S and S-PLUS*. Springer Verlag, New York.

Ringler, C., Bhaduri, A. & Lawford, R. (2013). The nexus across water, energy, land and food (WELF): potential for improved resource use efficiency? *Current Opinion in Environmental Sustainability* 5:617-624.

Ritchie H. & Roser M. (2017). CO<sub>2</sub> and other Greenhouse Gas Emissions. Disponível em: <https://ourworldindata.org/co2-and-other-greenhousegas-emissions/>.

Ritz, C. & Streibig, J. (2008) Nonlinear Regression with R. Springer Verlag, New York.

Santos, L. V., et al. (2016). Second-generation ethanol: the need is becoming a reality. *Industrial Biotechnology* 12: 40-57.

Solomon, S., K. Singh, S. Bhatnagar, V.K. Madan. 1988. Effect of applied ethephon on foliar enzymes of late-planted sugarcane. *Indian Journal of Sugarcane Technology* 5(1): 61–65.

## INFLUÊNCIA DA SUPLEMENTAÇÃO POR FLAVONOIDE NO CRESCIMENTO DE CLONES COMERCIAIS DE *E. urophylla* E *E. urograndis*

Data de aceite: 12/02/2020

### Nicholas Vinícius Silva

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Luciana Souto Mofatto

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Mariana Teixeira Rebouças

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Lucas Miguel de Carvalho

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Sheila Tiemi Nagamatsu

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Marcelo Falsarella Carazzolle

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

### Jorge Lepikson Neto

Instituto SENAI de Inovação em Biomassa – Três  
Lagoas, MS. Brasil

### Gonçalo Amarante Guimarães Pereira

Laboratório de Genômica e BioEnergia,  
Departamento de Genética e Evolução,  
Microbiologia e Imunologia, UNICAMP, Campinas,  
São Paulo, 13083-970, Brasil

**RESUMO:** Eucalyptus é um gênero nativo da Austrália que possui como característica a alta adaptabilidade e rápido crescimento em diversas condições edafoclimáticas. Atualmente o Eucalyptus é utilizado na produção de celulose e papel devido a essas características, além de apresentar alto potencial para a produção de biocombustível de segunda geração por ter uma produção significativa de biomassa. No entanto, a produção desta matéria prima é impactada por alguns fatores como a baixa disponibilidade de água e em casos extremos, à seca. Estudos de expressão gênica demonstraram que a suplementação com flavonoides modifica a estrutura da lignina, alterando a síntese de seus monômeros. Nesse contexto, este estudo teve como objetivo analisar se a suplementação com flavonoides em dois clones comerciais de eucalipto (*E. urophylla* e *E. urograndis*) está relacionada com o aumento de crescimento dessas plantas, tornando-a de maior interesse para a indústria energética. As análises foram



realizadas na plataforma R. Os resultados comprovaram que as plantas tratadas apresentaram um crescimento e diâmetro do caule superior às plantas controle em ambas espécies de eucalipto.

**PALAVRAS-CHAVE:** flavonoides, biomassa, *Eucalyptus*, chalcona, naringenina.

**ABSTRACT:** *Eucalyptus* is a genus native to Australia with high adaptability and rapid growth in many edaphoclimatic conditions. *Eucalyptus* is currently used for pulp and paper production due to these characteristics and has high potential for second generation biofuel production due to significant biomass production. However, the production of this raw material is impacted by some factors such as low water availability and in extreme cases, drought. Gene expression studies have shown that flavonoid supplementation modifies the structure of the line, altering the demonstration of its monomers. In this context, this study aimed to analyze the flavonoid supplementation in two commercial eucalyptus clones (*E. urophylla* and *E. urograndis*), which is useful with the growth of these plants, presenting a greater interest for the economic industry. The analyzes were performed on the R platform. The results proven that the treated plants presented superior growth and stem diameter than the control plants in eucalyptus species.

**KEYWORDS:** flavonoids, biomass, *Eucalyptus*, chalcone, naringenin.

## 1 | INTRODUÇÃO

*Eucalyptus*, é um gênero pertencente à família Myrtaceae, originário da Austrália que corresponde à aproximadamente setecentas espécies catalogadas (Longue Júnior e Colodette, 2013; Yang et al., 2016). O Eucalipto é plantado em mais de 90 países com mais de 18 milhões de hectares, sendo a árvore utilizada para a extração de madeira mais difundida do mundo (Arena et al., 2016; Suganya e Senthil Kumar, 2018). O plantio desta árvore têm se destacado devido ao rápido crescimento, alta produtividade, densidade de madeira relativamente alta, propriedades químicas apropriadas, baixo teor de umidade, facilidade de colheita e ampla adaptabilidade às características edafoclimáticas (Rockwood et al., 2008; Vallejos et al., 2017; Zhou et al., 2017), sendo plantados atualmente mais de cinco milhões de hectares de eucalipto no Brasil, com uma produção anual de madeira de 156,3 milhões de m<sup>3</sup> (www.cib.org.br). O gênero compreende cerca de 70% da produção florestal sustentável do país, sendo *Eucalyptus grandis* a espécie mais cultivada no Brasil, graças ao seu potencial produtivo e às características de sua madeira (Martins et al. 2007).

Nesse contexto, as espécies de eucalipto surgem com grande potencial para produção de biocombustíveis, no entanto, a madeira de eucalipto possui alta complexidade entre as moléculas de celulose, lignina e hemicelulose contidas na

parede celular. Esse é um fator limitante ao processo de aproveitamento de biomassa e obtenção de açúcares fermentescíveis para a produção de biocombustível, tornando o processo de deslignificação mais ineficiente (Pu et al., 2013). Além disso, existe uma grande variedade de resíduos advindos da indústria do papel que podem ser convertidos em combustíveis e fontes de carbono (carvão vegetal, biocarvão e carvão ativado) (Elyounssi et al., 2012; Kumar et al., 2010; Heidari et al., 2019). No caso da casca de eucalipto, ela apresenta-se importante principalmente na produção de calor e eletricidade por combustão (termoelétricas) e métodos utilizados em sistemas de bioenergia (biorefinarias), alguns métodos: Combined Heat and Power Plant (CHP), Heat Plant (HP), Combined heat and power with carbon capture and storage (BECCS) (Cavalett et al., 2018; da Silva et al., 2019); no entanto, a alta recalcitrância de sua estrutura lenhosa pode dificultar o processo de desconstrução da biomassa (Carvalho et al., 2015).

Apesar do eucalipto possuir alta capacidade em produzir biomassa (Longue Júnior e Colodette, 2013), fatores como a baixa disponibilidade de água, podem impactar negativamente o rendimento da produção (Chaves, 2002). Dessa forma, em busca de alternativas para a melhoria da madeira na produção de papel e celulose e seu aproveitamento para produção de biocombustíveis, nosso grupo avaliou a influência da suplementação de mudas de eucalipto com flavonoides.

É conhecido que a suplementação com os flavonoides narigenina e narigenina-chalcona em clones híbridos comerciais jovens de *Eucalyptus urograndis* resultou em alterações significativas da parede celular, reduzindo o teor e alterando a composição monomérica da lignina (monômeros siringil e guaiacil (S/G) ), além de aumentar a acessibilidade da parede celular a açúcares fermentescíveis (Lepikson-Neto et al., 2013; Salazar et al., 2013). A análise do transcriptoma do xilema das plantas tratadas revelou que as alterações são decorrentes da expressão diferencial de genes relacionados à lignificação e à resposta ao estresse abiótico, sugerindo que aplicação dos flavonoides poderiam afetar a resistência das plantas e o processo de formação da lignina (Lepikson-Neto et al., 2014).

Em estudos prévios, identificamos uma relação entre a suplementação com flavonoides e alterações na parede celular, bem como alterações na expressão de genes da via da lignina, associada ao crescimento e desenvolvimento vegetal. Por este motivo, avaliamos a aplicação do flavonoide narigenin-chalcona em plantas jovens de *Eucalyptus urophylla* e *E. urograndis* (híbrido entre *E. grandis* e *E. urophylla*) para verificar se existe um efeito sobre o crescimento de plantas jovens. Neste capítulo, iremos analisar os dados biológicos por meio de o software R e detalhar a aplicação de testes estatísticos.

## 2 I MATERIAIS E MÉTODOS

### 2.1 Preparação das plantas e suplementação

Mudas de clones comerciais de *E. urophylla* e *E. Urograndis*, fornecidas pela International Paper (Mogi-Guaçu, Brasil) com um mês de idade, foram transplantadas em vasos e mantidas em casa de vegetação do laboratório de Genética e Expressão (LGE) no Instituto de Biologia da Universidade Estadual de Campinas. As plantas foram submetidas à um período de adaptação por aproximadamente 30 dias com irrigação diária. Após esse período, as plantas foram divididas em dois grupos, controle (*E. urograndis*: C-UGR, *E. urophylla*: C-UPH) e suplementado (*E. urograndis*: T-UGR, *E. urophylla*: T-UPH) com flavonoide narigenin-chalcona. Todas as plântulas foram suplementadas semanalmente com 50-150 ml de solução nutritiva Sarruge (1975), e ao grupo suplementado, foi adicionado à solução o flavonoide narigenin-chalcona (0.01 mmol) (Lepkison-Neto et al., 2014), durante um período de seis meses. A altura das plantas foi mensurada semanalmente.

### 2.2 Organização dos dados

Os dados de parâmetros de crescimento para os grupos tratado e controle foram cadastrados em planilhas no MS Excel da seguinte maneira (para *E. urograndis* e *E. urophylla*):

- a) Plantas Controle: (1) altura (cm); (2) diâmetro do caule (cm);
- b) Plantas com tratamento: (1) Crescimento em altura (cm); (2) Diâmetro do caule (cm).

Desta maneira, as tabelas foram inseridas no MS Excel como abaixo:

Amostra	Semanas								
	W0	W1	W2	W3	W4	W5	W6	W...	W17
UGR_C1	4.47	5.19	5.37	6.84	7.00	7.21	7.23	...	9.27
UGR_C2	4.63	5.70	5.79	7.30	7.90	8.12	8.22	...	9.89
UGR_C3	6.60	8.20	8.83	11.1	12.5	12.8	12.98	...	13.2
UGR_C4	4.00	4.92	5.08	7.14	8.02	8.20	8.32	...	8.95
...	...	...	...	...	...	...	...	...	...
UGR_C10	5.61	6.46	6.46	8.27	9.02	9.13	9.28	...	10.44

Na primeira coluna estão os nomes das amostras (parcelas) e nas colunas 2 a 18 estão as medidas das plantas da semana inicial (W0) até semana final (W17). Cada planilha foi salva como “*Texto (separado por tabulações)*”, formato que permite importação dos dados para o R. Os dados de controle e tratamento foram salvos separadamente para facilitar a análise estatística e a composição dos gráficos.

## 2.3 Instalação de pacotes para gráficos

Para que os gráficos sejam gerados ao final das análises estatísticas, será utilizado o pacote “*ggplot2*”. Este pacote permite incluir atributos estéticos, deixando a visualização dos dados elegante (Wickham et al., 2016). É necessário usar o seguinte comando para instalar o pacote:

```
>install.packages("ggplot2")
```

Para utilização do pacote, digite:

```
>library("ggplot2")
```

## 2.4 Definindo o diretório de trabalho a ser usado nas análises dos dados

Antes de iniciar a análise é necessário definir o diretório de trabalho, especificando para o R em qual local estão os arquivos para importação dos dados. Para verificar o seu atual diretório de trabalho, digite:

```
>getwd()
```

Caso não seja o diretório em que você deseja trabalhar, será necessário definir o local através do comando:

```
>setwd("/diretorio desejado/")
```

## 2.5 Usando o R para importação e análise estatística dos dados

Como os dados foram salvos em arquivos texto separados por tabulações, é possível realizar a importação dos mesmos para o R através do comando:

```
# Tabela de crescimento (Controle)
```

```
>tb_ugr_cr_ctr <- read.delim(file = "UGR_crescimento_controle.txt", sep = "\t", header = TRUE, row.names = 1)
```

```
# Tabela de crescimento (Tratamento)
```

```
>tb_ugr_cr_trt <- read.delim(file = "UGR_crescimento_tratamento.txt", sep = "\t", header = TRUE, row.names = 1)
```

Os parâmetros da função *read.delim* indicam que: a) Os dados estão separados por tabulação (*sep = "\t"*); b) A tabela já possui cabeçalho (dados da primeira linha) (*header = TRUE*); c) Os nomes das linhas da tabela serão os mesmos que os dados da primeira coluna (*row.names = 1*).

Duas variáveis do tipo *Data Frame* foram criadas ao importar os dados (*tb\_ugr\_cr\_ctr* para Controle e *tb\_ugr\_cr\_trt* para Tratamento). *Data Frame* é uma estrutura de dados que permite ter colunas com tipos de dados diferentes (numéricos, caracteres, lógicos). Para obter as dimensões do *Data Frame*, utilize o comando *dim*:

```
# Número de linhas
```

```
>linhas_cr_ctr = dim(tb_ugr_cr_ctr)[1] # Controle
```

```
>linhas_cr_trt = dim(tb_ugr_cr_trt)[1] # Tratamento
```

```
# Número de colunas
```

```
>cols_cr_ctr = dim(tb_ugr_cr_ctr)[2] # Controle
>cols_cr_trt = dim(tb_ugr_cr_trt)[2] # Tratamento
```

## 2.6 Verificação da distribuição dos dados

Para verificar se a distribuição dos dados, foi utilizado o teste de Shapiro-Wilk. Para isto, variáveis do tipo vetor foram geradas para conter as medianas dos dados de cada semana (W0 até W17) por grupo, conforme código abaixo:

```
# Cálculo da mediana dos dados por semanas - Controle
>mediana_ugr_cr_ctr = vector(mode = "numeric", length = cols_cr_ctr)
>for(i in 1:cols_cr_ctr){
  mediana_ugr_cr_ctr[i] = median(tb_ugr_cr_ctr[,i], na.rm = TRUE)
}
# Teste Shapiro-Wilk
>shapiro.test(mediana_ugr_cr_ctr)
```

Os dados apresentaram uma distribuição normal com  $p\text{-value} > 0,05$ .

```
Shapiro-Wilk normality test
```

```
data: mediana_ugr_cr_ctr
```

```
W = 0.95342, p-value = 0.4811
```

Para o grupo de tratamento, as mesmas linhas de comando para o teste de Shapiro-Wilk foram repetidas. O resultado do teste está abaixo:

```
Shapiro-Wilk normality test
```

```
data: mediana_ugr_cr_trt
```

```
W = 0.94589, p-value = 0.3642
```

## 2.7 Análise estatística

Para identificar se houve efeito no tratamento das plantas por semana, foi utilizado um teste estatístico para a comparação das duas medianas dos grupos “Controle” e “Tratamento”. Para isto foi escolhido o *teste t de Student* para amostras independentes, conforme abaixo:

```
>t.test(tb_ugr_cr_ctr, tb_ugr_cr_trt,paired = FALSE)
```

Os resultados mostraram que houve diferença estatisticamente significativa ( $p\text{-value} < 0,05$ ), ou seja, o tratamento causou um aumento no crescimento das plantas.

```
Welch Two Sample t-test
```

```
data: tb_ugr_cr_ctr and tb_ugr_cr_trt
```

```
t = -6.1491, df = 334.44, p-value = 2.221e-09
```

```
alternative hypothesis: true difference in means is not equal to 0
```

95 percent confidence interval:

-11.110339 -5.724803

sample estimates:

mean of x mean of y

59.65465 68.07222

## 2.8 Geração de gráficos

Para melhor demonstração dos resultados, foram gerados gráficos de pontos e *boxplots* para demonstrar como o tratamento com flavonoide causou aumento no crescimento nas alturas das plantas. Para isto, foi utilizado o pacote *ggplot2*:

```
>library(ggplot2)
```

Para facilitar a plotagem dos dados, as variáveis *tempo*, *crescimento* e *grupos* (tratado e controle) foram criadas a partir dos cálculos da mediana, como demonstrado abaixo:

```
>tempo = colnames(tb_ugr_cr_ctr)
>crescimento_ctr = mediana_ugr_cr_ctr
>crescimento_trt = mediana_ugr_cr_trt
>crescimento = c(crescimento_ctr,crescimento_trt)
>grupos = c(rep("C_UGR", length(crescimento_ctr)), rep("T_UGR", length(crescimento_trt)))
>tb_crescimento = data.frame(crescimento,grupos,tempo)
```

A variável *tempo* é um vetor utilizado no *eixo x* do gráfico de pontos, a qual contém os nomes das semanas (obtido pelo comando *colnames* retorna os nomes das colunas do cabeçalho do *Data Frame*).

```
> tempo
[1] "W0" "W1" "W2" "W3" "W4" "W5" "W6" "W7" "W8" "W9" "W10" "W11" "W12" "W13" "W14"
"W15" "W16" "W17"
```

A variável *crescimento* também é um vetor que possui valores das medianas dos grupos Controle e Tratamento.

```
> crescimento
[1] 43.75 45.00 46.00 47.25 50.40 55.00 55.20 56.85 57.50 57.75 60.00 61.00 63.00 66.00 68.00
72.00 75.00 75.50
[19] 47.00 48.50 49.25 53.00 56.10 59.75 60.75 64.20 65.55 66.35 71.40 72.00 77.50 80.50 83.00
86.00 88.50 90.00
```

A variável *grupos* também é um vetor contendo a especificação dos grupos Controle e Tratamento.

```
> grupos
[1] "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_
UGR" "C_UGR" "C_UGR" "C_UGR"
[14] "C_UGR" "C_UGR" "C_UGR" "C_UGR" "C_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR"
"T_UGR" "T_UGR" "T_UGR"
```



```
[27] "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR" "T_UGR"
```

Ao final, foi gerada a variável *tb\_crescimento* do tipo *Data Frame* contendo os dados de tempo, crescimento e grupos.

```
> head(tb_crescimento, 5)
```

	crescimento	grupos	tempo
1	43.75	C_UGR	W0
2	45.00	C_UGR	W1
3	46.00	C_UGR	W2
4	47.25	C_UGR	W3
5	50.40	C_UGR	W4

Após a formatação dos dados das medianas, o gráfico de pontos foi gerado e atribuído à variável *p1* conforme o código abaixo. Foi utilizado o comando *tiff* para salvar o gráfico automaticamente (figura 1).

```
## Gráfico de pontos para comparação do crescimento usando os dados das medianas
>p1 <- ggplot(data = tb_crescimento, aes(x = factor(tempo, levels = unique(tempo)), y = crescimento))
+
  geom_point(aes(colour = factor(grupos)), size = 3) +
  labs(x = "Tempo (Semanas)", y = "Crescimento (Altura)", colour = "Grupos") + theme_gray(base_size
= 14)
>tiff(filename = "grafico_cresc_UGR_pontos_mediana.tif", width = 800, height = 600, units = "px",
pointsize = 12, compression = "lzw+p", bg = "white", res = NA, family = "", restoreConsole = TRUE,
type = c("windows", "cairo"))
>p1
>dev.off()
```

Para os demais gráficos (figuras 2 e 3), os dados foram formatados conforme

**Nota 1** para gerar a tabela abaixo:

```
> tb_ugr_cr = read.table(file = "UGR_crescimento_formatado.txt", header = TRUE, row.names = 1)
> head(tb_ugr_cr, 5)
```

	valores_cr	grupos_cr	semanas_cr
1	47.5	C_UGR	W0
2	45.5	C_UGR	W0
3	54.5	C_UGR	W0
4	42.2	C_UGR	W0
5	41.2	C_UGR	W0

Para gerar os gráficos de pontos (figura 2) e do tipo *Boxplot* (figura 3) foi utilizado o código abaixo. Todos os gráficos são salvos automaticamente em formato *tiff*.

```
## Data Frame com os dados formatados
>tb_ugr_cr = >read.table(file = "UGR_crescimento_formatado.txt", header = TRUE, row.names = 1)
## Gráfico de pontos para comparação do crescimento usando todos dados
>p2 <- ggplot(subset(tb_ugr_cr, !is.na(valores_cr)), aes(x = factor(semanas_cr, levels =
unique(semanas_cr)), y = valores_cr)) +
```

```

geom_point(aes(colour = factor(grupos_cr)), size = 3) +
labs(x = "Tempo (semanas)", y = "Crescimento (Altura)", colour = "Grupos")+ theme_gray(base_size =
14)
>tiff(filename = "grafico_cresc_UGR_pontos_totais.tif", width = 800, height = 600, units = "px", pointsize
= 12, compression = "lzw+p", bg = "white", res = NA, family = "", restoreConsole = TRUE, type =
c("windows", "cairo"))
>p2
>dev.off()
## Gráfico Boxplot para comparação do crescimento usando todos os dados
>p3 <- ggplot(data = subset(tb_ugr_cr, !is.na(valores_cr)), aes(x = factor(semanas_cr, levels =
unique(semanas_cr)), y = valores_cr)) +
  geom_boxplot(aes(fill = grupos_cr), position=position_dodge(0.9)) +
  scale_fill_manual(values = c("Red", "Blue")) +
  labs(x = "Tempo (semanas)", y = "Crescimento (Altura)", fill = "Grupos") + theme_gray(base_size =
14)
>tiff(filename = "grafico_cresc_UGR_boxplot.tif", width = 800, height = 600, units = "px", pointsize
= 12, compression = "lzw+p", bg = "white", res = NA, family = "", restoreConsole = TRUE, type =
c("windows", "cairo"))
>p3
>dev.off()

```

### 3 | DISCUSSÃO DOS DADOS BIOLÓGICOS

Os flavonoides são compostos polifenólicos encontrados em todas as plantas vasculares e não vasculares. Embora não sejam essenciais para o crescimento e desenvolvimento das plantas, os flavonoides têm papéis específicos da espécie na nodulação, fertilidade, defesa e proteção contra radiação ultravioleta (UV). Foi demonstrado que os flavonoides modulam o transporte de auxina, além das respostas trópicas dependentes da auxina. No entanto, os flavonoides não são reguladores essenciais desses processos porque o transporte e as respostas trópicas ocorrem na sua ausência. Os flavonoides modulam a atividade das glicoproteínas P transportadoras de auxinas e parecem modular a atividade de proteínas reguladoras, como fosfatases e quinases. A análise filogenética sugere que os mecanismos de transporte de auxinas evoluíram na presença de compostos flavonoides produzidos para a eliminação de espécies reativas de oxigênio e defesa contra herbívoros e patógenos (Bessau et al., 2007; Peer e Murphy, 2007).

A aplicação em plantas jovens de dois clones comerciais de eucalipto durante seis meses, promoveu o crescimento das plantas e o teste estatístico confirmou que houve diferença estatisticamente significativa ( $p\text{-value} < 0,05$ ). No entanto, há relatos de que o acúmulo de flavonoides resulta na redução do crescimento devido à

diminuição do fluxo de auxina em um mutante de *A. thaliana* (Bessau et al, 2007). Foi relatado que o flavonoide naringenin inibe a atividade do 4CL *in vitro* (Voo et al. 1995), suprime o crescimento e reduz o conteúdo de lignina em plantas gramíneas, como milho (*Zea mays* L.), Arroz (*Oryza sativa* L.) e capim-água (*Echinochloa oryzicola*) (Deng et al. 1998). O efeito da suplementação de naringenin e chalcona suprimiram o crescimento de 20 espécies anuais de plantas, embora as sensibilidades foram diferentes entre as plantas (Chen et al., 2004).

Em eucalipto, a suplementação de plantas jovens com flavonoides promoveu a modificação da madeira, como resultado de alterações na composição dos monômeros que formam a lignina. Com a diminuição desses monômeros (do tipo S e G), as árvores suplementadas possuem a madeira mais susceptível à degradação química, além de apresentar alterações no metabolismo de açúcar e de genes responsivos à stress (Lepikson-Neto et al., 2014). Nesse contexto, os resultados apresentados demonstram uma nova aplicação das chalconas como moléculas que regulam o crescimento de plantas jovens de eucalipto, possivelmente pelo desbalanço ocasionado na via dos fenilpropanoides.

#### 4 | NOTA 1

Código para formatação dos dados para uso em gráfico do pacote “ggplot2”:

```
### Grupos
>grupos_cr = c(rep("C_UGR",linhas_cr_ctr*cols_cr_ctr),
               rep("T_UGR",linhas_cr_trt*cols_cr_trt))
### Tempo (Semanas) - pelo nome das colunas do grupo Controle
>nomes_col_cr_ctr = colnames(tb_ugr_cr_ctr)
>semanas_cr_ctr = vector(mode="character", length = >length(nomes_col_cr_ctr)*linhas_cr_ctr)
>for(i in 1:length(nomes_col_cr_ctr)){
  temp = rep(nomes_col_cr_ctr[i],linhas_cr_ctr)
  if(i == 1){
    semanas_cr_ctr = temp
  }
  else{
    semanas_cr_ctr = c(semanas_cr_ctr,temp)
  }
}
### Tempo (Semanas) - pelo nome das colunas do grupo Tratamento
>nomes_col_cr_trt = colnames(tb_ugr_cr_trt)
>semanas_cr_trt = vector(mode="character", length = >length(nomes_col_cr_trt)*linhas_cr_trt)
>for(i in 1:length(nomes_col_cr_trt)){
  temp = rep(nomes_col_cr_trt[i],linhas_cr_trt)
  if(i == 1){
    semanas_cr_trt = temp
```

```

}
else{
  semanas_cr_trt = c(semanas_cr_trt,temp)
}
}
>semanas_cr = c(semanas_cr_ctr,semanas_cr_trt)
## Valores de crescimento - Controle
>valores_cr_ctr = vector(mode="numeric",length = linhas_cr_ctr*cols_cr_ctr)
>for(i in 1:cols_cr_ctr){
  temp = tb_ugr_cr_ctr[,i]
  if(i == 1){
    valores_cr_ctr = temp
  }
  else{
    valores_cr_ctr = c(valores_cr_ctr,temp)
  }
}
## Valores de crescimento - Tratamento
>valores_cr_trt = vector(mode="numeric",length = linhas_cr_trt*cols_cr_trt)
>for(i in 1:cols_cr_trt){
  temp = tb_ugr_cr_trt[,i]
  if(i == 1){
    valores_cr_trt = temp
  }
  else{
    valores_cr_trt = c(valores_cr_trt,temp)
  }
}
>valores_cr = c(valores_cr_ctr, valores_cr_trt)
### Tabela final formatada ###
>tabela_final_cr = cbind(valores_cr,grupos_cr,semanas_cr)
### Salvar a tabela formatada ###
>write.table(tabela_final_cr, file = "UGR_crescimento_formatado.txt", sep = "\t", row.names = TRUE,
col.names = TRUE)

```

## 5 | FIGURAS

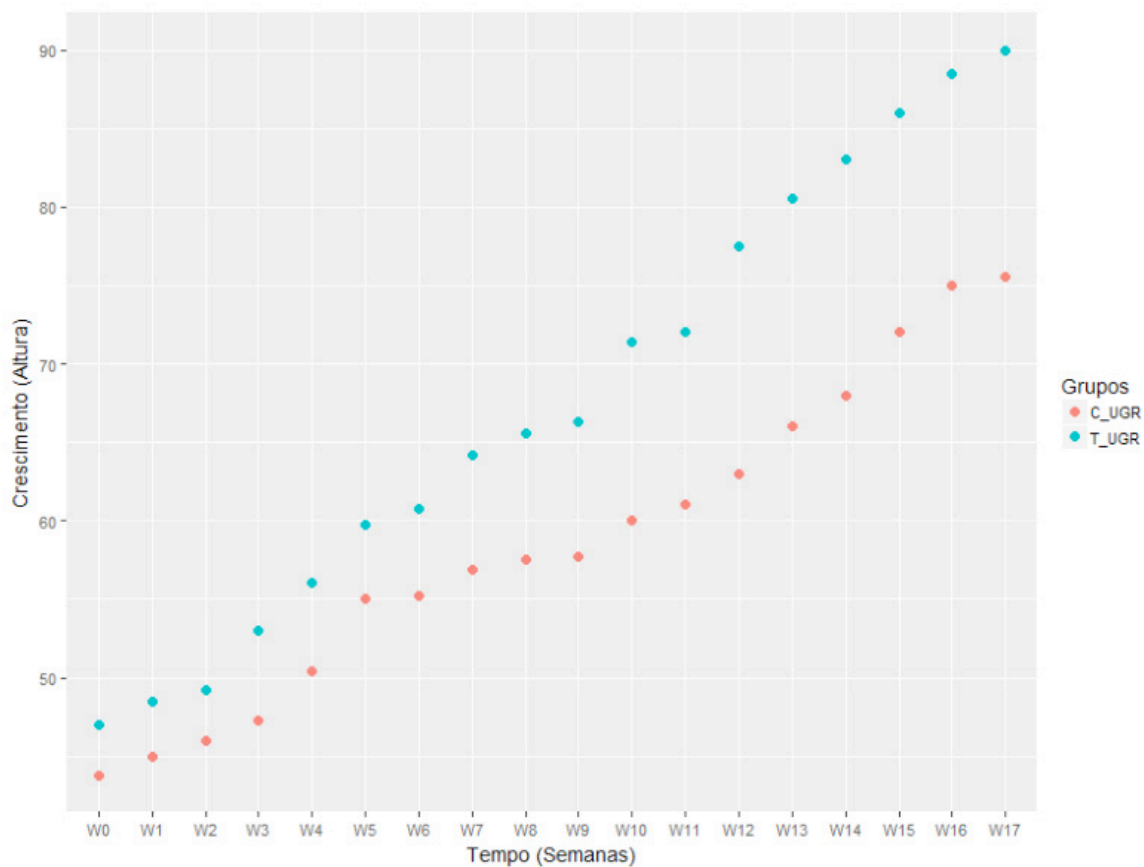


Figura 1. Gráfico de pontos (Crescimento em função do tempo em semanas) referente a medianas das medidas em *E. urograndis* (C\_UGR = Controle e T\_UGR = Tratamento).

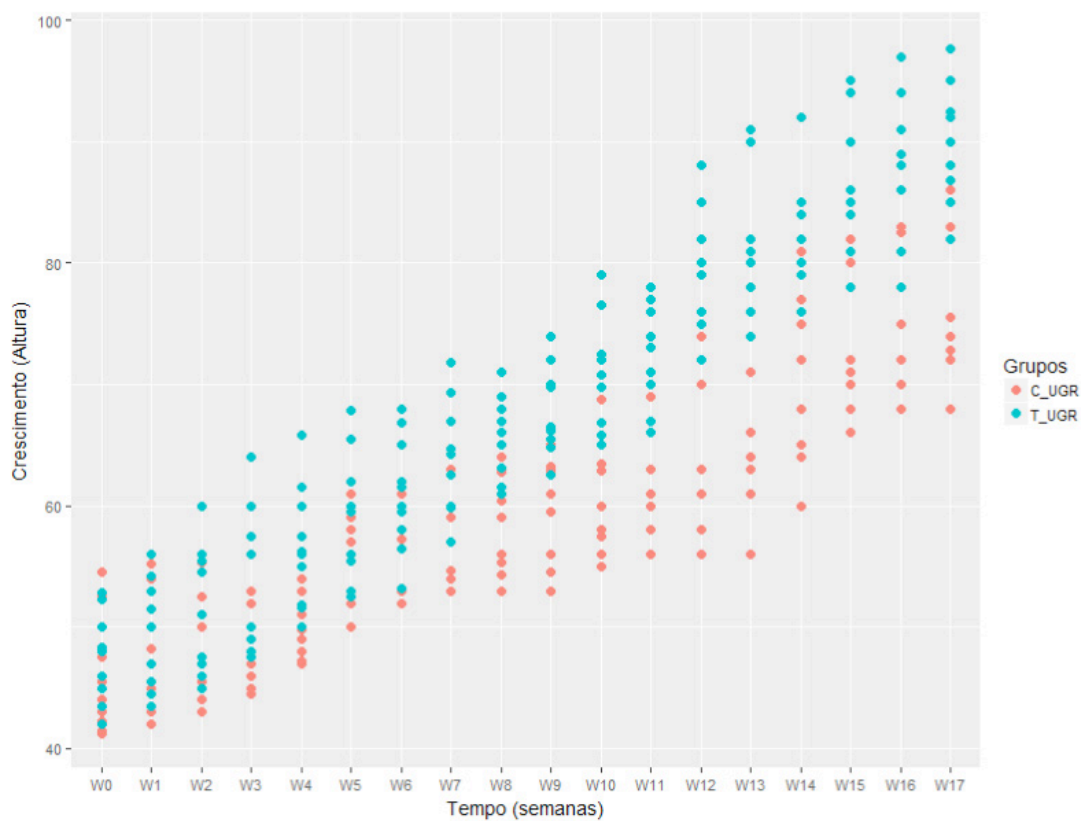


Figura 2. Gráfico de pontos (Crescimento em função do tempo em semanas) referente a distribuição das medidas de crescimento em *E. urograndis* (C\_UGR = Controle e T\_UGR = Tratamento).

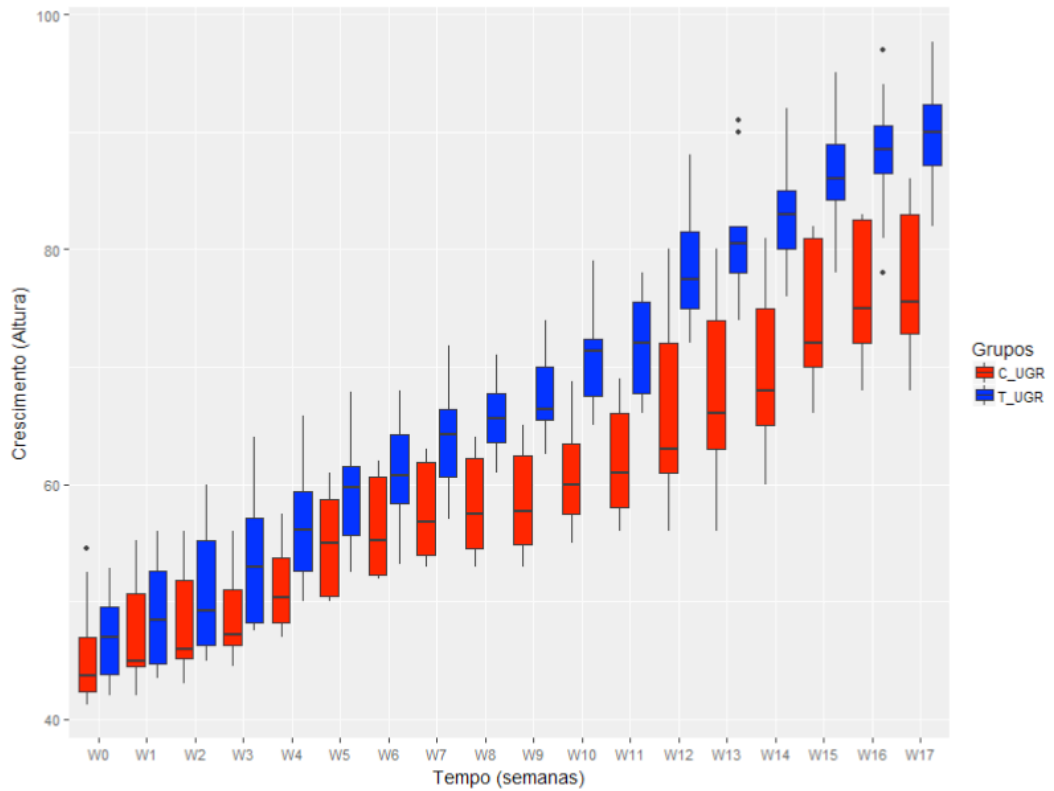


Figura 3. *Boxplot* (Crescimento em função do tempo em semanas) referente a todos os dados de crescimento em *E. urograndis* (C\_UGR = Controle e T\_UGR = Tratamento).

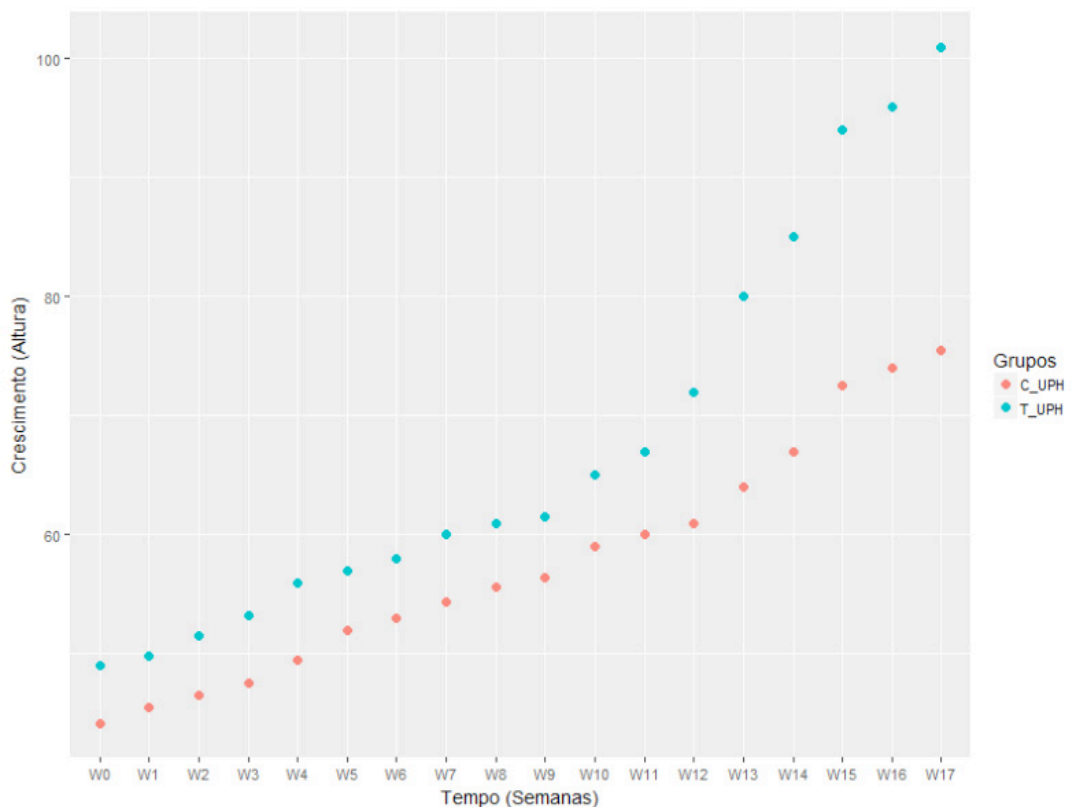


Figura 4. Gráfico de pontos (Crescimento em função do tempo em semanas) referente a mediana das medidas de crescimento em *E. urophylla* (C\_UPH = Controle e T\_UPH = Tratamento).



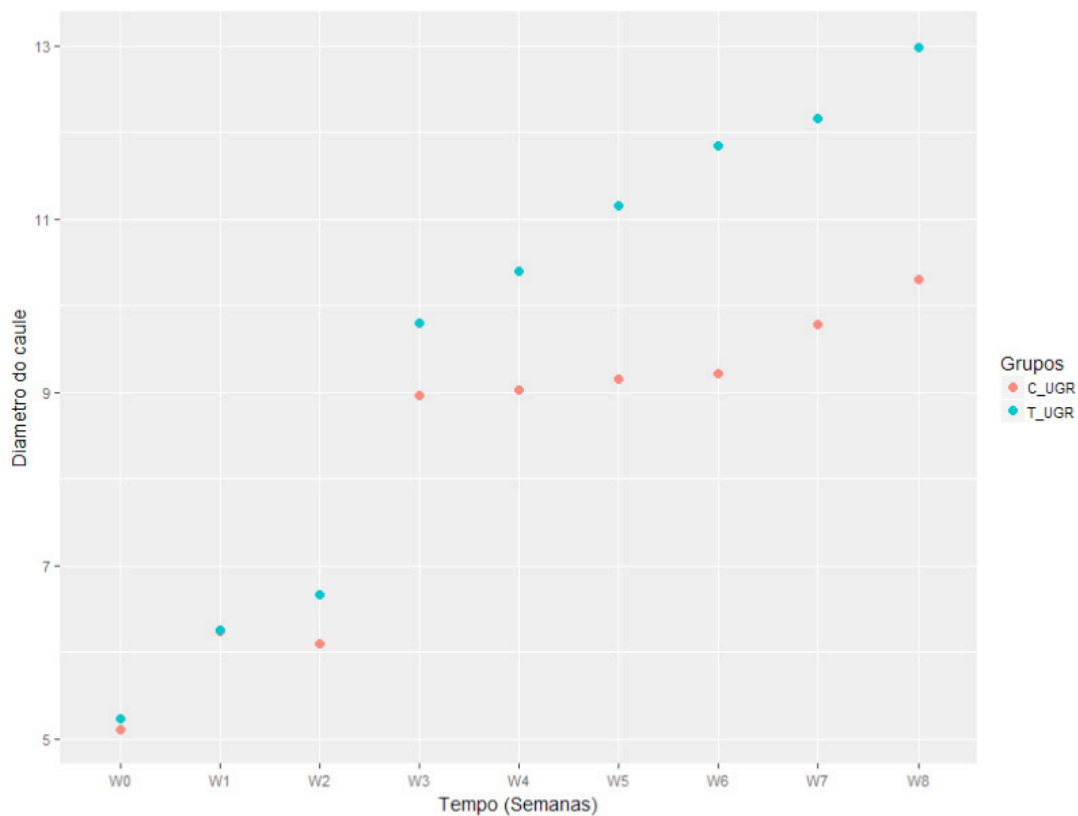


Figura 5. Gráfico de pontos (Diâmetro do caule em função do tempo em semanas) referente a mediana das medidas de diâmetro em *E. urograndis* (C\_UGR = Controle e T\_UGR = Tratamento).

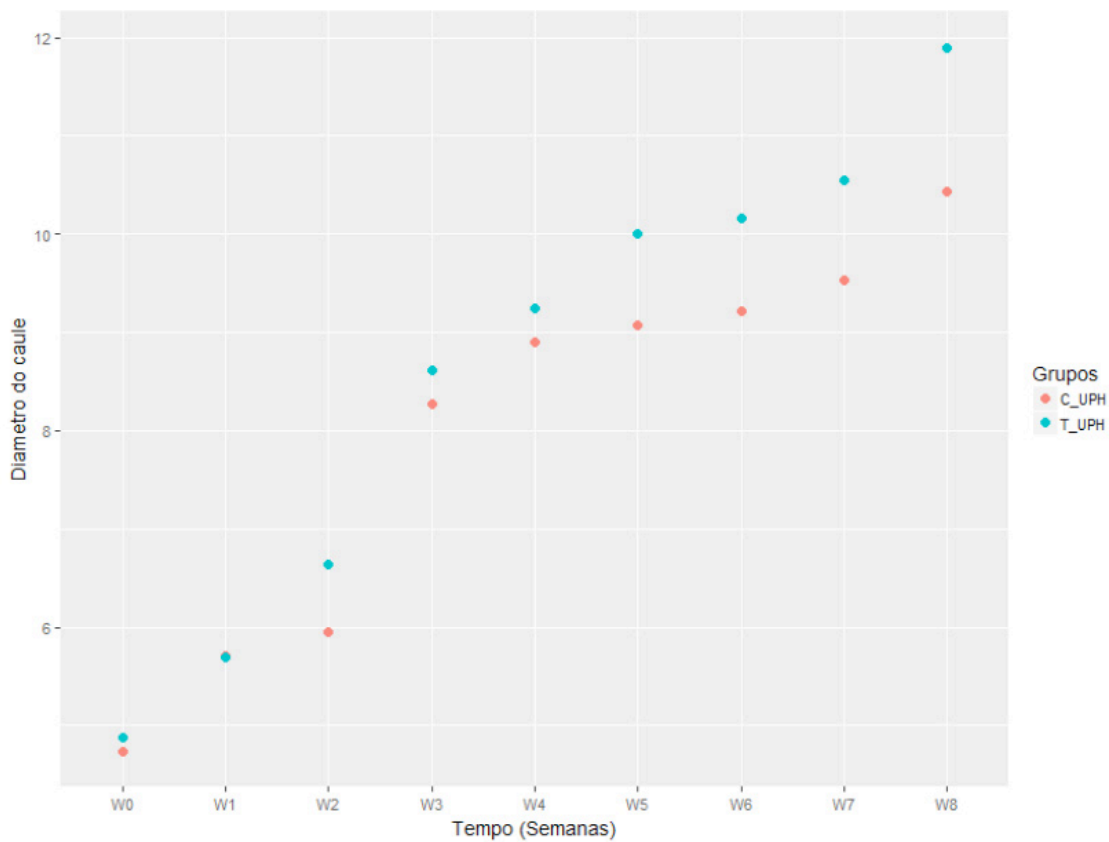


Figura 6. Gráfico de pontos (Diâmetro do caule em função do tempo em semanas) referente a mediana das medidas de diâmetro do caule em *E. urophylla*.

## 6 | ACESSIBILIDADE AOS DADOS E COMANDOS (SCRIPT)

Link: [doi.org/10.13140/RG.2.2.33780.09600](https://doi.org/10.13140/RG.2.2.33780.09600)

### AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ (Processo: 150977/2019-0; 140869/2016-6) e a Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (Processo: 2016/10312-8) pelo auxílio financeiro à pesquisa. Ao Centro de Computação em Engenharia e Ciências - FAPESP/Cepid (2013/08293-7). À empresa International Paper do Brasil, pela disponibilização do material genético (clones).

### REFERÊNCIAS

Arena, N.; Lee, J.; Clift, R.; (2016) Life Cycle Assessment of activated carbon production from coconut shells. *J. Clean. Prod.* 125, 68e77.

Bessau, S.; Hoffmann, L.; Geoffroy P.; et al. (2007) Flavonoid Accumulation in Arabidopsis Repressed in Lignin Synthesis Affects Auxin Transport and Plant Growth. *Plant Cell Online*, v. 19, n. 1, p. 148-162.

Cavalett, O.; Norem Slettmo, S.; Cherubini, F. (2018) Energy and environmental aspects of using eucalyptus from Brazil for energy and transportation services in Europe. *Sustainability*, 10, 4068.

Carvalho, D.M.; Sevastyanova, O.; Penna, L.S.; et al. (2015) Assessment of chemical transformations in eucalyptus, sugarcane bagasse and straw during hydrothermal, dilute acid, and alkaline pretreatments. *Ind. Crops Prod.* 73, 118-126.

Chaves, M.M. (2002) How Plants Cope with Water Stress in the Field? *Photosynthesis and Growth. Annals of Botany*, v. 89, n. 7, p. 907-916.

Chen, W. et al (2004) Effects of root-applied naringenin and chalcone on the growth of annual plants. *Weed Biol Monag*, v. 238, P. 235-238.

da Silva, C.M.S; Vital,B.V.; Rodrigues, F.A.; de Almeida, E.W.; Carneiro,A.C.O.; Cândido, W.L. (2019) Hydrothermal and organic-chemical treatments of eucalyptus biomass for industrial purposes, *Bioresource Technology*, v. 289.

Deng, F.; Aoki, M.; Yogo, Y. (1998) Effect of naringenin on the monolignol biosynthesis and the growth of annual plant. *J. Weed Sci. Technol.* 43 (Suppl.), 292-293 (in Japanese).

Elyounssi, K.; Collard, F.X.; Mateke, J.-a.N.; Blin, J. (2012) Improvement of charcoal yield by two-step pyrolysis on eucalyptus wood: a thermogravimetric study. *Fuel* 96, 161e167.

Heidari, A.; Khaki, E.; Younesi, H.; Lu, R.H. (2019) Evaluation of fast and slow pyrolysis methods for bio-oil and activated carbon production from eucalyptus wastes using a life cycle assessment approach, *Journal of Cleaner Production*, v. 241.

Kumar, G.; Panda, A.K.; Singh, R.K. (2010) Optimization of process for the production of bio-oil from eucalyptus wood. *J. Fuel Chem. Technol.* 38 (2), 162e167.

Lepikson-Neto, J.; Alves, A.; Simões, R.; Deckmann, A.C.; Camargo, E.L.O.; Salazar, M.M.; Rio,

- M.C.S. do; Nascimento, L.C.; Pereira, G.A.G.; Rodrigues, J.C. (2013) Flavonoid supplementation reduces the extractive content and increases the syringyl/guaiacyl ratio in *Eucalyptus grandis* x *Eucalyptus urophylla* hybrid trees. *Bioresources* (Raleigh, N.C), v. 8(2):1747-1757.
- Lepikson-Neto, J.; Nascimento, L.C.; Salazar, M. M.; Camargo, E.L.O.; Cairo, J.P.F.; Teixeira, P.J.P.L.; Marques, W.L.; Squina, F.M.; Mieczkowski, P.; Deckmann, A.C.; Pereira, G.A.G. (2014) Flavonoid supplementation affects the expression of genes involved in cell wall formation and lignification metabolism and increases sugar content and saccharification in the fast-growing eucalyptus hybrid *E. urophylla* x *E. grandis*. *BMC Plant Biology*, 14:301.
- Longue Júnior, D.; Colodette, J.L. (2013) Importância e versatilidade da madeira de eucalipto para a indústria de base florestal. *Pesquisa Florestal Brasileira*, v. 33, n. 76, p. 429-438.
- Martins, A.; Cardoso, A.L.; Stahl, J.A.; Diniz, J. (2007) Low temperature conversion of rice husks, eucalyptus sawdust and peach stones for the production of carbon-like adsorbent. *Bioresour Technol* 198:1095–1100.
- Peer, W.A.; Murphy, AS (2007) Flavonoids and auxin transport: modulators or regulators? *Trends in Plant Science* 12, 556–563
- Petersen, M.; Strack, D.; Matern, U. (2010) Biosynthesis of Phenylpropanoids and Related Compounds. *Annual Plant Review*, v. 40, p. 182-257.
- Pu, Y. et al., (2013) Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. *Biotechnology for biofuels*, v. 6, n. 1, p. 15.
- Rockwood, D. L. et al. (2008) Energy product options for Eucalyptus species grown as short rotation woody crops. *International Journal of Molecular Sciences*, 9,1361–1378.
- Salazar, M.M.; Nascimento, L.C.; Camargo, E.L.O.; Gonçalves, D.; Lepikson-Neto, J.; Marques, W.L.; Teixeira, P.J.P.L.; Mieczkowski, P.; Mondego, J.M.C.; Carazzolle, M.F.; Deckmann, A.C.; Pereira, G.A.G. (2013) Xylem transcription profiles indicate potential metabolic responses for economically relevant characteristics of Eucalyptus species. *BMC Genomics*, 14:201.
- Sarruje, J. (1975) Soluções Nutritivas. *Summa Phytopyologica*, v.1, p. 231-233.
- Suganya, S.; Senthil Kumar, P.; (2018) Evaluation of environmental aspects of brew waste-based carbon production and its disposal scenario. *J. Clean. Prod.* 202, 244e252.
- Vallejos, M.E.; Kruyeniski, J.; Area, M.C. (2017) Second-generation bioethanol from industrial wood waste of South American species. *Biofuel. Res. J.* 4 (3), 654e667.
- Voo, K.S.; Whetten, R.W.; O'Malley, D.M.; Sederoff, R.R. (1995) 4-Coumarate: coenzyme A ligase from loblolly pine xylem. Isolation, Characterization, and complementary DNA cloning. *Plant Physiol.* 108, 85–97.
- Wang, W.; Vinocur, B.; Shoseyov, O.; Altaman, A. (2004) Role of Plant Heat Shock Proteins and molecular chaperons in the abiotic stress response. *Trends in Plant Science*, v. 9, p. 244-252.
- Wickham H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Yang, X.; Lii, D.; Mcgrouter, K.; Long, W.; Li, Y.; Chen, Y.; Lv, X.; Niazi, N.K.; Song, Z.; Wang, W. (2016) Effect of Eucalyptus forests on understory vegetation and soil quality. *International symposium on forest soils. Journal of Soils Sediments*.
- Zhou, H.; Qian, Y.; Kraslawski, A.; Yang, Q.; Yang, S. (2017) Life-cycle assessment of alternative liquid fuels production in China. *Energy* 139, 507e522.

## **SOBRE OS ORGANIZADORES**

**Écio Souza Diniz** - Pós-doutor em Engenharia Florestal pela Universidade Federal de Viçosa – UFV (2018-presente) e pesquisador colaborador do Departamento de Engenharia Florestal da UFV, vinculado ao Laboratório de Geoprocessamento e Sensoriamento Remoto. Consultor em análises estatísticas e processamento de dados e diretor executivo da empresa Beta Analítica. Biólogo pelo Centro Universitário de Lavras – UNILAVRAS (2009), mestre em Engenharia Florestal pela Universidade Federal de Lavras – UFLA (2013), doutor em Botânica pela Universidade Federal de Viçosa – UFV (2017). Pesquisador convidado em estágio doutoral de colaboração internacional no Institute of Landscape Ecology da University of Münster (Alemanha) e no AgroEcoSystem Research Department da University Bayreuth (Alemanha) em 2015. Tem experiência e conhecimento nas linhas de pesquisa em ecologia funcional e filogenética com foco em Florestas Neotropicais, fitossociologia, invasão biológica por plantas, fragmentação florestal, monitoramento florestal de longo-prazo (dinâmica), modelagem climática para espécies vegetais, geoprocessamento e sensoriamento remoto, modelagem estatística de dados de vegetação e análises de dados via linguagem R. Possui afinidade principalmente pela ecologia e evolução de ambientes de vegetação de Cerrado e Mata Atlântica, além de conhecimentos significativos de Campos rupestres. Atua atualmente como pesquisador conduzindo estudos e participando de projetos nas áreas citadas, revisor de periódicos científicos e fornecedor de consultorias, cursos e treinamentos em análises de dados e estatística. E-mail: [eciodiniz@gmail.com](mailto:eciodiniz@gmail.com), <https://orcid.org/0000-0002-3543-6571>, [https://www.researchgate.net/profile/Ecio\\_Souza\\_Diniz](https://www.researchgate.net/profile/Ecio_Souza_Diniz), <http://lattes.cnpq.br/4486399846318757>, <https://betaanalitica.com.br/>

**Pedro Manuel Villa** - Engenheiro Agrônomo (2002), estudante de Biologia (2000-2004), Mestre em Ecologia Tropical (2007) pela Universidad de Los Andes (ULA), Doutor em Botânica (2018) pela Universidade Federal de Viçosa. Foi pesquisador associado do Instituto Nacional de Investigações Agrícolas (INIA-Mérida, 2005-2006). Foi pesquisador coordenador da área Agroflorestal do Instituto Nacional de Investigações Agrícolas (INIA-Amazonas, 2008-2014). Realizou estágio profissional (2009) no Laboratório de Plant Production Systems da Universidade de Wageningen, Holanda. Foi assessor de projetos ambientais em comunidades indígenas da Amazônia com o Fundo Mundial para o Meio Ambiente (GEF) – Programa das Nações Unidas para Desenvolvimento (PNUD) (2010- 2015). Tem experiência em pesquisa na área de ecologia de comunidades e ecossistemas. Atualmente é revisor de periódicos nacionais e internacionais como Ecological Indicators, Forest Ecology and Management, Agroforestry systems, Revista de Biologia Tropical, Acta Botanica Brasílica. Atualmente é vinculado como pós-doutor do LARF – Laboratório de Restauração Florestal do Departamento de Engenharia Florestal da Universidade Federal de Viçosa – UFV. E-mail: [villautana@gmail.com](mailto:villautana@gmail.com), <https://orcid.org/0000-0003-4826-3187>, [https://www.researchgate.net/profile/Pedro\\_Villa](https://www.researchgate.net/profile/Pedro_Villa), <http://lattes.cnpq.br/2943038330953932>

## ÍNDICE REMISSIVO

### A

Abundance 3, 14, 16, 17, 20, 21, 22, 23, 24, 25, 26, 27, 69, 72, 74, 75, 79, 80, 97, 110

Abundância 21, 70, 96, 101, 102, 103, 104, 106, 107, 110

Análise de correlação 43

Análise multivariada 43, 45, 55

*Arabidopsis thaliana* 122, 123, 125, 126, 135, 136

### B

Biomassa 57, 83, 99, 106, 138, 139, 140, 150, 151, 152

Bootstrapping 12, 15

### C

Cana-de-açúcar 81, 138, 139, 140, 141, 142, 146, 147, 148

Cana energia 138, 140, 141, 142, 143, 144, 146, 147, 148

*Caryocar brasiliense* 110, 111, 121

Chalcona 151, 152, 153, 159

Chemical variability 110

Classificação de habitats 56

Cobertura vegetal 21, 44

Computational programming 1

Crescimento 12, 36, 107, 125, 138, 139, 140, 141, 142, 143, 144, 147, 148, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162

### D

Dados da vegetação 1, 2

Desmatamento 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 80

Dominance 12, 13, 14, 16, 18, 20, 21, 23, 24, 27, 97

Dominância 12, 21, 96, 98, 106, 107

Dominância de árvores 12

### E

Ecologia do fogo 96

Elevação 56, 58, 59, 60, 61, 63, 64, 65, 66, 67

Estatística 2, 30, 31, 34, 38, 42, 43, 45, 53, 59, 142, 144, 145, 147, 153, 154, 155, 166

Estrutura da vegetação 21, 70, 96

Estrutura florestal 56

Ethephon 138, 139, 140, 141, 142, 143, 145, 146, 147, 148, 149

*Eucalyptus* 150, 151, 152, 164, 165

Evapotranspiração 82, 83, 85, 86, 89, 90, 93, 94

Expressão diferencial 122, 124, 126, 127, 128, 133, 134, 135, 152

Extrapolação com base em amostras 12

## F

Fatores bióticos e abióticos 56, 57  
Flavonoids 112, 113, 114, 151, 165  
Floresta secundaria 12  
Forest planting 70  
Forest regrowing 12  
Forest restoration 13, 29, 70, 79

## G

Geostatistics 70, 71, 74  
Gradiente ambiental 43, 53, 56

## I

Importance value index 20, 27, 28  
Índice de valor de importância 21, 102  
Insects 110, 111, 112

## K

*Kriging* 56, 57, 69, 73, 75

## M

Modelagem matemática 139

## N

Naringenina 151

## P

Pacote agrewater 82, 83, 89, 90, 93  
Phytosociological characterization 70  
Programação computacional 2  
Propriedades do solo 43, 45, 46, 53

## R

R. Análise exploratória 30  
Rarefação 12  
Regeneração florestal 12  
Regeneração natural 96, 98, 107  
Resiliência 96, 98, 106, 107  
R language 1, 9, 22, 28, 123

## S

Safer 82, 83, 85, 86, 88, 90  
Sampled-based rarefaction and extrapolation 12



Savanização de florestas 96  
Second-growth forests 12, 13  
Shiny 30, 31, 32, 33, 37, 42, 133  
Soil attributes 44, 55, 70, 71, 74  
Soil nutrients 110, 111, 112, 113, 116, 117, 118, 119, 120  
Spatial variation 110, 113, 116, 117, 119, 121  
Statistics 1, 8, 74, 80, 120

## T

Transcriptômica de plantas 122  
Tree dominance 12

## V

Vegetation cover 7, 20  
Vegetation data 1, 3, 8, 28  
Vegetation structure 20, 21, 70

 **Atena**  
Editora

**2 0 2 0**