UNIVERSIDADE FEDERAL DO PARANÁ

FRANCIEL EDUARDO REX

# SPATIAL CHARACTERIZATION OF WOODY SPECIES DIVERSITY IN TROPICAL SAVANNAS USING GEDI AND OPTICAL DATA,

CURITIBA

2024

FRANCIEL EDUARDO REX

# SPATIAL CHARACTERIZATION OF WOODY SPECIES DIVERSITY IN TROPICAL SAVANNAS USING GEDI AND OPTICAL DATA

Tese apresentada ao Programa de Pós-Graduação em Engenharia Florestal, Setor de Ciências Agrárias, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Engenharia Florestal, Área de Concentração: Manejo Florestal.

Orientadora: Prof. Ana Paula Dalla Corte

Coorientadores: Dr. Carlos Alberto Silva
Drª. Carine Klauberg Silva

CURITIBA

2024

Ficha catalográfica elaborada pela
Biblioteca de Ciências Florestais e da Madeira - UFPR

Rex, Franciel Eduardo
Spatial characterization of woody species diversity in Cerrado (Brazillian Savanna) using multiple remote sensing data sources / Franciel Eduardo Rex.- Curitiba, 2024.
1 recurso on-line : PDF

Orientadora: Profª. Drª. Ana Paula Dalla Corte
Coorientadores: Dr. Carlos Alberto Silva
Drª. Carine Klauberg Silva
Tese (Doutorado) - Universidade Federal do Paraná, Setor de Ciências Agrárias. Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 10/04/2024.

1. Cerrados - Brasil. 2. Cerrados - Sensoriamento remoto. 3. Plantas dos cerrados - Brasil. 4. Plantas lenhosas - Brasil. 5. Biodiversidade florestal - Brasil. 6. Monitorização ambiental - Brasil. I. Dalla Corte, Ana Paula. II. Silva, Carlos Alberto. III. Silva,Carine Klauberg. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.

CDD - 577.30981
CDU - 581.526.424(81)
634.0.182.3(81)
528.8

Bibliotecária: Marilene do Rocio Veiga - CRB9/424

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS AGRÁRIAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA FLORESTAL - 40001016015P0

# TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **FRANCIEL EDUARDO REX** intitulada: **SPATIAL CHARACTERIZATION OF WOODY SPECIES DIVERSITY IN TROPICAL SAVANNAS USING GEDI AND OPTICAL DATA**, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 10 de Abril de 2024.

Assinatura Eletrônica
12/04/2024 11:07:53.0
ALEXANDRE BEHLING
Presidente da Banca Examinadora

Assinatura Eletrônica
11/04/2024 17:52:12.0
VERALDO LIESENBERG
Avaliador Externo (UNIVERSIDADE DO ESTADO DE SANTA CATARINA)

Assinatura Eletrônica
11/04/2024 14:52:50.0
VAGNER ALEX PESCK
Avaliador Externo (UNIVERSIDADE ESTADUAL DO CENTRO-OESTE)

Assinatura Eletrônica
11/04/2024 18:09:39.0
ANA PAULA DALLA CORTE
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
12/04/2024 23:08:23.0
CAIO HAMAMURA
Avaliador Externo (INSTITUTO FEDERAL DE EDUCAçãO, CIêNCIA E TECNOLOGIA DO ESTADO DE SãO PAULO)

“Dedico esse trabalho aos meus pais Rosani e Dilmar Rex, e também aos meus amados avós Olga e Roque Ziembowicz.

# AGRADECIMENTOS

Aos meus pais Rosani e Dilmar, por serem minha base, pelo apoio incondicional em todos os momentos, e por acreditarem nos meus sonhos. Amo muito vocês.

Aos meus avós Olga e Roque, pelo carinho, confiança e força que sempre me transmitiram. Obrigado também por entenderem que eu precisaria mais uma vez estar longe de casa.

À minha Orientadora Prof.ª Dr.ª Ana Paula Dalla Corte, que me abriu portas desde o início do mestrado, me orientando desde sempre para ir em busca do melhor. Agradeço a orientação, dedicação, confiança em mim e por todos os ensinamentos durante o curso.

Meus sinceros agradecimentos aos pesquisadores e coorientadores deste trabalho, Carine Silva e Carlos Alberto Silva por todo suporte, esforços e contribuições para a melhoria desta pesquisa, e pelo seu tempo e disposição em sempre ajudar.

As minhas amigas de longa data que me acompanharam nesta jornada, em especial a Pâmela, Bruna, Franciele, Fernanda, Iaci. Obrigado por dividirem este processo e por me escutar em momentos de dificuldade.

Aos meus amigos do Crossfit, em especial a Manu que virou uma grande amiga e parceira de treino. Agradeço por deixar os dias mais leves, tranquilos e descontraídos. Nada como jogar uns pesos para o alto.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de estudos.

Ao Programa de Pós-graduação em Engenharia Florestal da UFPR.

À Universidade Federal do Paraná (UFPR) por todas as oportunidades desde o mestrado.

“Somos o que fazemos, mas somos,
principalmente, o que fazemos para
mudar o que somos.”
Eduardo Galeano

**RESUMO**

O desenvolvimento da capacidade de monitorizar a diversidade de espécies, em todo o mundo, é de grande importância para travar a perda de biodiversidade. Para este fim, a detecção remota desempenha um papel único. Neste estudo, avaliamos o potencial dos dados de *Global Ecosystem Dynamics Investigation* (GEDI) combinados com imagens ópticas de satélite convencionais e dados de reanálise climática para prever a diversidade alfa *in situ* (riqueza de espécies, índice de Simpson e índice de Shannon-Wiener) de espécies arbóreas. A integração de dados auxiliares pode melhorar as previsões de métricas de biodiversidade no cerrado (savana tropical brasileira). Dados de imagens ópticas do satélite Sentinel-2, dados climáticos do ERA 5, imagens SRTM-DEM e dados GEDI simulados foram selecionados para a caracterização da diversidade em nossas quatro áreas de estudo. No total, 50 parcelas de 900m² foram mensuradas em campo para calibrar modelos de regressão de floresta aleatória (RF) com o objetivo de estimar índices de diversidade de espécies arbóreas a partir de variáveis de sensoriamento remoto. Selecionamos previamente variáveis de cada conjunto de dados para compor os modelos de índices de diversidade. Para tal, as seguintes variáveis foram selecionadas: GEDI (RH98 + FHD + PAI + COV); Sentinel-2 (NDVI + SR + SAVI + RDVI + EVI); SRTM (Slope + Aspect + Elevation) e ERA-5 (Temperature + Precipitation). Na plataforma Google Earth Engine (GEE), processamos as variáveis de sensoriamento remoto com exceção dos dados GEDI que foram simulados no R Studio. Em seguida integramos todas as variáveis nos footprints dos dados GEDI para aplicar os modelos desenvolvidos para toda a cobertura do bioma Cerrado e assim obtermos a caracterização espacial de cada índice. Por fim, aplicamos uma análise de incerteza nos dados estimados. Modelos de regressão de floresta aleatória (RF) foram adequados para estimar índices de diversidade de espécies arbóreas a partir de variáveis de sensoriamento remoto. A partir desses modelos, geramos mapas de índice de diversidade para todo o cerrado utilizando todos os dados do GEDI disponíveis em órbita. Para todos os modelos, a diversidade métrica estrutural da altura da folhagem (FHD) foi selecionada, e o índice de vegetação por diferença renormalizada (RDVI) também foi selecionado em todos os modelos de diversidade de espécies. Para o modelo de Shannon foram selecionadas duas variáveis GEDI. No geral, os modelos tiveram desempenho para diversidade de espécies variando de ($r^2$= 0,24 a 0,56). Em termos de RMSE%, o modelo de Shannon apresentou o menor valor entre os índices de diversidade (31,98%). Nossos resultados sugerem que os modelos desenvolvidos são ferramentas valiosas para avaliar a diversidade de espécies em ecossistemas de savana tropical, embora cada modelo possa ser escolhido com base nos objetivos de um determinado estudo, na quantidade alvo de desempenho/erro e na disponibilidade de dados. Além disso, a contribuição dos produtos desenvolvidos pode servir principalmente para fomentar políticas públicas no sentido de buscar a proteção dos recursos naturais do bioma em questão.

**Palavras-chave:** Cerrado, diversidade alpha, GEDI, LiDAR, imageamento, modelagem.

**ABSTRACT**

Developing the capacity to monitor species diversity worldwide is of great importance to halt biodiversity loss. To this end, remote sensing plays a unique role. In this study, we evaluate the potential of Global Ecosystem Dynamics Investigation (GEDI) data combined with conventional satellite optical imagery, and climate reanalysis data to predict in situ alpha-diversity (Species richness, Simpson index, and Shannon–Wiener index) of tree species. The integration of ancillary data can improve biodiversity metrics predictions in Cerrado (Brazilian Tropical Savanna). Data from Sentinel-2 optical imagery, ERA-5 climate data, SRTM-DEM imagery, and simulated GEDI data were selected for the characterization of diversity in our four study areas. In total, 50 plots of 900m² were measured in the field to calibrate random forest (RF) regression models with the objective of estimating tree species diversity indices based on remote sensing variables. We previously selected variables from each data set to compose the diversity index models. To this end, the following variables were selected: GEDI (RH98 + DHF + PAI + COV); Sentinel-2 (NDVI + SR + SAVI + RDVI + EVI); SRTM (Slope + Aspect + Elevation) and ERA-5 (Temperature + Precipitation). On the Google Earth Engine (GEE) platform, we processed the remote sensing variables except for the GEDI data that were simulated in R Studio. Then we integrated all variables into the GEDI data footprints to apply the models developed to the entire biome and thus obtain the spatial characterization of each index. Finally, we apply an uncertainty analysis to the estimated data. Random Forest (RF) regression models were suitable for estimating tree species diversity indices from remote sensing variables. From these models, we generated diversity index maps for the entire Cerrado using all GEDI data available in orbit. For all models, the structural metric Foliage Height Diversity (FHD) was selected, and the Renormalized Difference Vegetation Index (RDVI) was also selected in all species diversity models. For Shannon's model, two GEDI variables were selected. Overall, the models had performance for species diversity ranging from ($R^2$= 0.24 to 0.56). In terms of RMSE%, the Shannon model had the lowest value among the diversity indices (31.98%). Our results suggested that the developed models are valuable tools for assessing species diversity in tropical savanna ecosystems, although each model can be chosen based on the objectives of a given study, the target amount of performance/error, and the availability of data. Furthermore, the contribution of the products developed can serve mainly to promote public policies to seek the protection of natural resources in the biome.

**Keywords**: Cerrado, Alpha diversity, GEDI, LiDAR, Imagery, Modeling

**LIST OF FIGURES**

**LIST OF TABLES**

**SUMMARY**

## 1 INTRODUCTION

The Cerrado Bioma, also known as Brazilian Savanna, size of 2M square kilometers and harbors more than 12,400 plant species (Damasco et al., 2018) and contains one-third of Brazilian biodiversity with a high number of endemisms, making it the most biodiverse savanna on the planet (Klink and Machado, 2005). Cerrado is considered the largest savanna area in South America and is recognized as the continent’s second-largest ecosystem after the Amazon (Klink and Machado, 2005). However, this ecosystem has been severely degraded and impacted by anthropic action. By 2022, about half of the Cerrado area had already been converted to pasture (51.6%) or cultivation area (26%) (MMA, 2022), while only a small portion (8.2%) of the biome is formally protected by indigenous parks or reserves (BRASIL, 2016). Historically, most of the efforts to estimate forest cover changes in South America have been focused on tropical rainforests, with far less attention dedicated to the regions with shorter humid seasons (Pennington et al., 2006, Portillo-Quintero and Sánchez-Azofeifa, 2010). Because of vegetation cover changes, tree species biodiversity declined substantially (Khare et al., 2019), resulting in the loss of biodiversity and important ecosystem services, such as nutrient and water cycle regulation, soil protection, as well as food and wood provision (Costanza et al., 2014; Milheiras and Mace, 2019).

Spatial monitoring of ecosystem structure and diversity indices estimates are very important information, therefore new approaches are needed, since this type of information can adequately guide conservation policies and actions (Valbuena et al., 2020), in addition to serving as support for climate change mitigation strategies (Hurtt et al., 2019).

Ecologists generally conduct floristic inventories through intensive field sampling to preserve and better understand species diversity in each area. However, for large areas, estimating species diversity from field surveys has several disadvantages, mainly related to time, expense, intrinsic difficulties developing standardized procedures for collecting reproducible data and assuring temporal assessment (Palmer et al., 2002). To overcome these difficulties, species diversity monitoring systems would identify the location, magnitude, and rate of vegetation change, thus informing effective conservation strategies and management plans to prevent or mitigate biodiversity losses and increase resilience capacity of forest areas

(Mallinis et al., 2020). To that end, remote sensing can provide a valuable source of information for understanding the structure and composition of an ecosystem at scales relevant to wide-area biodiversity monitoring (Gillespie et al., 2008; Laurin et al., 2016). When compared to conventional field-based approaches to investigate species, remote sensing technologies may offer a practical and cost-effective means of mapping, especially at broad scales (Heinzel and Koch, 2012).

Several remote sensing studies have assessed species diversity in savanna ecosystems (Hernández-Stefanoniet al., 2012; Arekhi et al., 2017; Mandosela et al., 2017; de Souza Mendes et al., 2019; Chrysafis et al., 2020). However, remote sensing studies for monitoring Cerrado biodiversity are still scarce and restricted to the use of conventional passive data, mainly in the savanna and grassland formations where there is low biomass and thus low signal saturation (de Souza Mendes et at., 2019). In addition, most studies that used multispectral information explored only the Normalized Difference Vegetation Index (NDVI) to assess tree species diversity (Hernández-Stefanoniet al., 2012; Madonsela et al., 2017), though other broadband VIs have been more rarely used (Khare et al., 2021). Despite being widely used in vegetation and remote sensing studies, the ability of NDVI saturation affects the discrimination of vegetation structure and composition as leaf area index peaks (Huete et al., 2002). Yet, conventional passive optical imagery from air- and space-borne platforms have limited applicability, as these data are not directly sensitive to vertical vegetation structure (Marselis et al., 2018). To satisfy this requirement, active remote sensing, especially using light detection and ranging (LiDAR) instruments, has clearly demonstrated the ability to accurately measure both spatial and vertical vegetation structure (Lim et al., 2003; Rex et al., 2020; Valbuena et al., 2020). Some studies have also shown accurate results using airborne LiDAR data for assessing species diversity (Laurin et al., 2014; Schäfer et al., 2016; Oldeland et al., 2010; Mallinis et al., 2020; Peng et al., 2018). However, most of these studies used airborne LiDAR data, which is limited by high survey costs and low area coverage.

A new era of Earth observation has shown great potential to surmount these challenges, with spaceborne platforms becoming integral for carbon accounting, hydrology, global-scale climate modeling, and biodiversity mapping (Almeida et al., 2020; Hakkenberg et al., 2023). Studies have shown great potential of satellite-derived structural information to assess diversity in large areas. Marselis et al., (2019) used GEDI structural metrics such as canopy height and plant area volume density and were able to explain up to 71% of variance in species diversity indexes (Shannon and Richness) in African savanna ecosystems. Although such methods have been gradually developed in recent decades and have shown advantages to highlight the role of forest diversity, it is not yet clear whether such methods could be applied in Brazilian savannas. Especially because some studies found that biodiversity-productivity relationships vary among tropical, temperate, and boreal forests and among different regions (e.g., Paquette and Messier, 2011; Liang et al., 2016. Crockett et al., 2023).

The National Aeronautics and Space Administration (NASA) launched the Global Ecosystem Dynamics Investigation (GEDI) spaceborne LiDAR sensor on December 5, 2018, as the first spaceborne sensor specifically designed to map 3D terrestrial vegetation structure (Dubayah et al., 2020). GEDI began collecting scientific data on March 25, 2019, thus opening a new era for large-scale investigations on biogeography and macroecology. GEDI is mounted on the International Space Station (ISS) in low Earth orbit between 51.6° north and south latitudes. GEDI contains three lasers, emitting 1064 nm pulses at 242 Hz that illuminate the Earth's surface with footprints of ~25 m diameter on the ground. Two of the lasers are full power, and one is split into two beams, producing a total of four beams, which are optically dithered to produce eight ground tracks; four full-power and four coverage beam tracks, with footprints separated by ~60 m along-track and 600 m across track (Dubayah et al., 2020). The GEDI mission, initially planned to last for two years, collected four years of data from April 2019 to March 2023. The instrument is currently stored on the International Space Station and is expected to resume collecting data in fall 2024. The goal of the GEDI mission is to provide data for studying forest structure and biomass in tropical and temperate zones (Dubayah et al., 2020). Hence, GEDI has a high potential to provide unprecedented opportunities for ecological and biodiversity studies (Valbuena et al., 2020; Crockett et al., 2023).

To the best of our knowledge, no study has yet proposed to assess the combined contribution of variables from different sensors, such as spectral vegetation indices (VIs), canopy structure, topography and climate data. Factors such as altitude, aspect, precipitation, and temperature play a decisive role in the distribution of vegetation communities (Jin et al., 2009), yet have rarely been considered to assess species diversity.

# 2 OBJECTIVES

## 2.1 MAIN OBJECTIVE

The main objective of this doctoral dissertation was to develop and evaluate multi-source predictive models (GEDI + optical, climate, and topographic data) to estimate three diversity indices (Shannon, Simpson, and Species Richness) in the Cerrado Biome.

## 2.2 SPECIFIC OBJECTIVES

- To understand how selected metrics derived from simulated GEDI, vegetation indices from multispectral imagery, climate and digital elevation models over four representative and selected study areas are related to three diversity indices.

- To characterize large-scale species diversity across the entire Cerrado (i.e., 1.9 million km²) by applying the calibrated Random Forest (RF) models to multi-source data across its entire extent.

- Aggregating the footprint level, three species diversity indices estimate to 1-km-resolution grid across the biome.

- Measure the uncertainties of RF model predictions for a grid with 1-km resolution grid across the entire biome.

- propose a categorization to classify species diversity into low, medium and high based on the distribution of estimated values for the entire biome.

- Propose strategies of data usage and discuss opportunities for future research.

# 3 HYPOTHESIS

The hypothesis of this work supports the idea that providing multiple sources of remote sensing can contribute to generating predictive models to obtain better estimates of the diversity of tree species in the Brazilian Cerrado. Topographic, climate and optical data associated with structural metrics (GEDI) can generate more robust models providing better estimates of the diversity of tree species in the Cerrado.

# 4 LITERATURE REVIEW

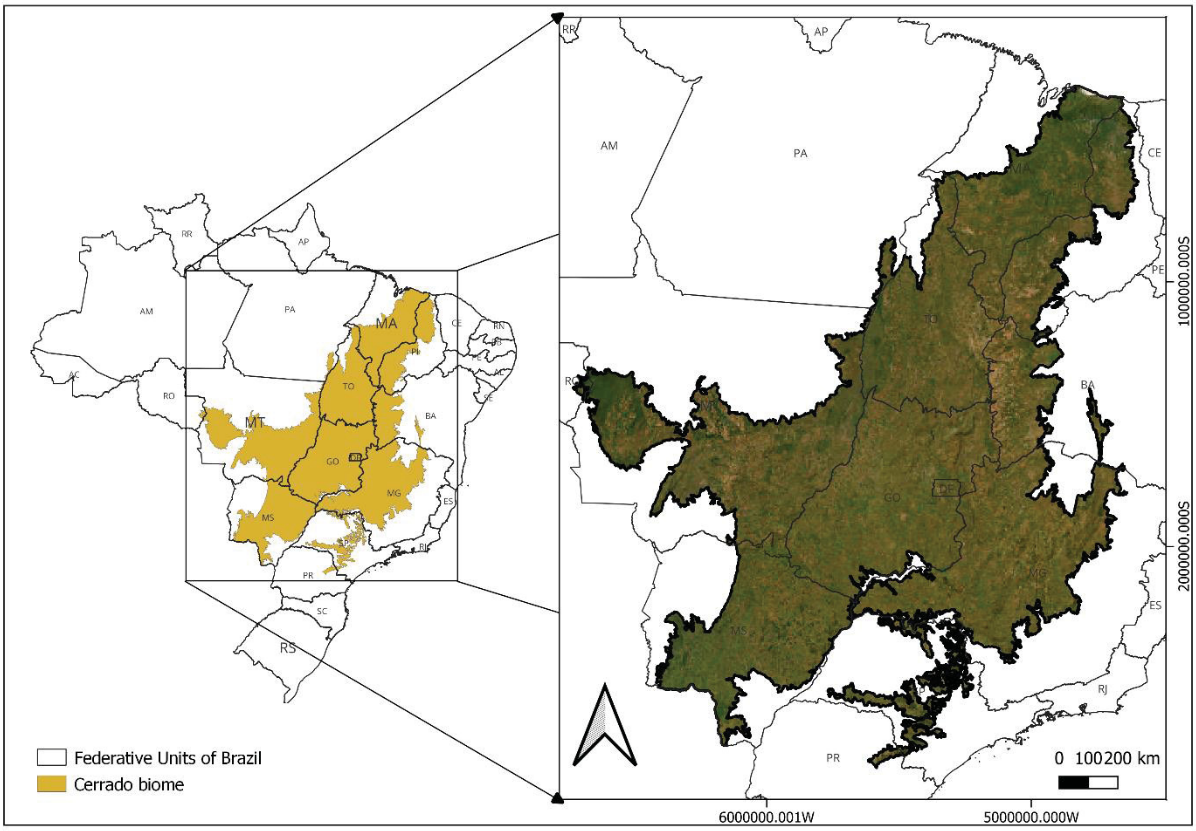
## 4.1 THE CERRADO BIOME

The Cerrado is the second largest Brazilian biome, only surpassed by the Amazon. It occupies around 21% of the national territory and is considered the last agricultural frontier on the planet (Borlaug, 2002). The Cerrado covers as a continuous area the States of Goiás, Tocantins and the Federal District, part of the states of Bahia, Ceará, Maranhão, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Piauí, Rondônia and São Paulo (FIGURE 1), and also It occurs in disjoint areas to the north in the States of Amapá, Amazonas, Pará and Roraima, and to the south, on small “islands” in Paraná state (EMBRAPA, 2023). A third of Brazilian biodiversity and around 5% of the world's flora and fauna are found in the Cerrado. It is considered the most biologically diverse savannah in the world (MMA, 1999; Sawyer, 2002).

The term Cerrado is commonly used to designate the set of ecosystems (savannas, forest formations, and grassland) that occur in Central Brazil (Ribeiro et al., 1981). In a physiognomic sense, the term savanna refers to areas with trees and shrubs spread over a grassy layer, without the formation of a continuous canopy. Forest formations represent areas with a predominance of tree species, where there is continuous or discontinuous canopy formation. The term grassland designates areas with a predominance of herbaceous species and some shrubs, with no trees in the landscape (EMBRAPA, 2023). These formations involve eleven main types of vegetation (FIGURE 2), divided into forest formations (Mata Riparian, Mata de Galeria, Mata Seca and Cerradão), savanna (Cerrado in the strict sense, Parque Cerrado, Palmeiral and Vereda) and grassland (Campo Sujo, Campo Limpo and

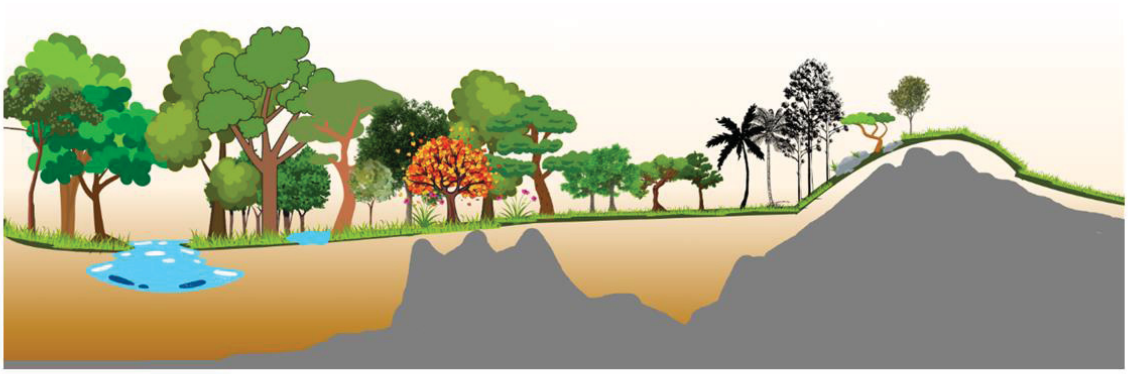
Campo Rupestre); they also involve subtypes, totaling up to 25 phytophysiognomies (Ribeiro and Walter, 2008).

FIGURE 1 - LOCATION OF THE CERRADO BIOME IN BRAZIL. (SOURCE: IBGE AND ESRI)




The diversity of phytophysiognomies in the Cerrado biome is remarkable and fundamental to understanding its complexity (Santos et al., 2018). Each of the formations (savannas, forest formations, and grasslands) have specific characteristics in terms of flora, fauna, topography and water regime (Cerqueira et al., 2023). Furthermore, the vegetation cover of the Brazilian Cerrado is extremely diverse, reflecting the ecological richness of this biome (Bendini et al., 2020). Composed of a variety of phytophysiognomies, this vegetation adapts to the different climatic, soil and rainfall conditions found in the region.

FIGURE 2 - CERRADO MAIN PHYTOPHYSIOGNOMIES. (ADAPTED FROM: EMBRAPA, 2023).




The Cerrado's physiognomic gradient is related to environmental factors and is maintained by a regime of disturbances, natural or anthropogenic, making the mosaic dynamic in time and space (Durigan et al., 2018). In general, the grassland formations are characterized by the predominance of herbaceous species and some shrubs, with the absence of trees. Savannas refer to areas with trees and shrubs spread over a layer of grasses, with the absence of continuous canopies. Forest formations are areas with a predominance of tree species, which may have a continuous or discontinuous canopy (Ribeiro and Walter 2008). When talking about the vegetation cover of the Cerrado, it is important to address these differences between phytophysiognomies (EMBRAPA, 2023):

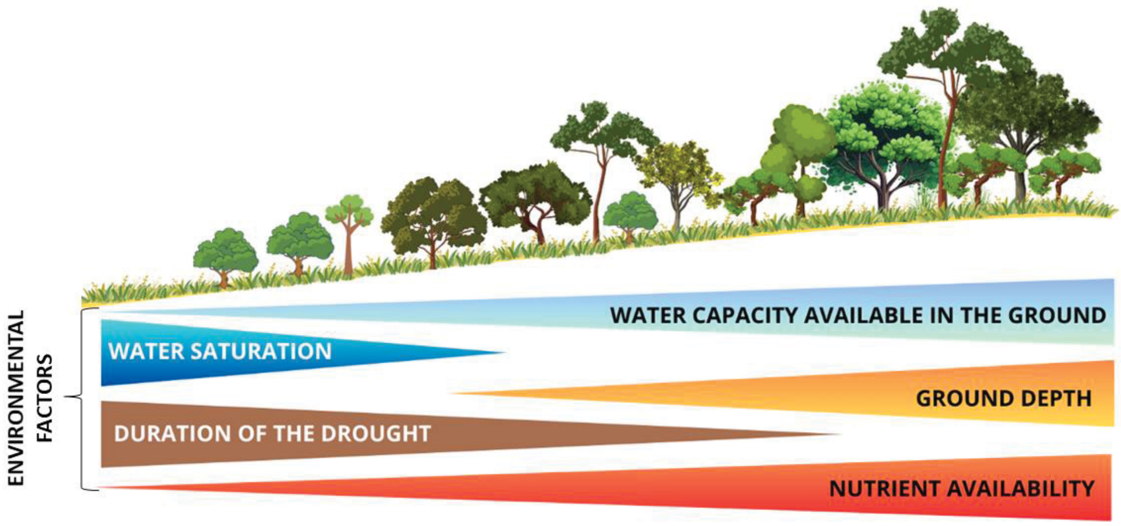
• The savanna formation, considered as Cerrado proper, is characterized by predominantly low vegetation, composed of grasses, shrubs and sparse trees, often adapted to seasonal fire. Like the Cerrado itself, the Cerrado park has more spaced trees, forming a "park" scenario between the plant groups. Areas with a marked presence of palm trees, such as buriti, generally associated with humid areas and bodies of water, form Veredas (EMBRAPA, 2023):

• Forest formations, such as gallery forests, have denser and taller vegetation, found along rivers and streams, generally with different tree species adapted to the constant water regime. The riparian forest, like the gallery forest, but with lower density and adapted to the banks of rivers and streams. The dry forest, on the other hand, is a lower density forest that occurs in areas with drier soil and presents specific adaptations to water scarcity. The Cerradão is a denser and taller variation of the Cerrado, with taller and denser trees than the Cerrado itself (Ribeiro and Walter, 1998).

• The grassland formation can be characterized as dirty, clean and rocky countryside (Ribeiro and Walter, 1998). The dirty field is low vegetation with the presence of grasses and a few trees or shrubs. The clean field is an area of low vegetation with few trees, generally with sandier and less fertile soil. Campo rupestre is a vegetation adapted to rocky terrain, with the presence of small trees, shrubs and species adapted to rockier soils (Pinheiro and Durigan, 2012).

These differences in phytophysiognomies reflect not only the plant species variety, but also adaptations to climate, soil and water availability in different parts of the Cerrado biome. This physiognomic gradient increases in its structure and plant biomass with increasing nutrient availability, soil depth and soil water content (Durigan et al., 2018) (FIGURE 3).

FIGURE 3 - SCHEMATIC REPRESENTATION OF ENVIRONMENTAL FACTORS RELATED TO VEGETATION STRUCTURE IN CERRADO REGIONS (ADAPTED FROM: DURIGAN et al., 2018).




The Cerrado is characterized by the presence of dry winters and rainy summers, a climate classified predominantly as Köppen's Aw (rainy tropical). It has an average annual rainfall of around 1,500 mm, varying from 750 mm to 2,000 mm (Adámoli et al., 1987). The rains occur concentrated from October to March (rainy season), and the average temperature of the coldest month is above 18 °C. The contrast between the lower surfaces (less than 300 m), the long plateaus between 900 m and 1,600 m and the extensive distribution in latitude give the Cerrado a very large thermal diversification. On the other hand, the general atmospheric mechanism determines a similar seasonal pattern of precipitation throughout the region, creating a tendency towards uniformity in rainfall (Nimer, 1989), with a dry season and a well-defined rainy season. To the south of the biome, in areas with a milder climate, the Cwa climate may occur (Eiten, 1994), which also characterizes the highest places in the central region, above 1,200 m altitude.

The wide range of vegetation formations is an essential part of the Cerrado's ecological biodiversity, housing a wide diversity of plant and animal species, some of which are exclusive to this environment (Santos et al., 2018; Cerqueira et al., 2023). Although the Cerrado has great ecological importance, until recently it was considered unproductive land, being used only for the extraction of charcoal, firewood and extensive livestock farming activities (Durigan et al., 2011). As a result, 55% of the Cerrado was deforested or transformed by this anthropization, making its territory highly fragmented (Machado et al., 2004). Due to this high degree of threat and the high number of endemic species, the Cerrado is considered one of the 35 global hotspots with priority for conservation (Mittermeier et al., 2011).

## 4.2 SPECIES DIVERSITY

According to Reis (2012), species diversity indicates variety of species, and may or may not address information about how species are distributed in one or more communities of living beings. The information that can be addressed is species richness and species evenness. Species richness refers to the number of species in each geographic area, region, or community. The greater the number of species, the greater the richness of a community. Equability is the relative abundance in each Community, which means that the closer the abundance of species within a community is, the greater the equability.

Several tools help researchers measure diversity, in particular species richness, richness estimators and diversity indices stand out (Magurran, 2004). Species diversity analysis aims to establish references that allow evaluating how diverse a forest formation is in terms of species (Calixto Júnior, 2009). In this research, we used three main indices, which are:

### 4.2.1 SPECIES RICHNESS

Species richness (i.e., number of species per unit area) is the simplest and most widely used indicator to describe communities and regional diversity (Guralnick et al., 2007; Feest et al., 2010; Magurran, 2011).

### 4.2.2 SHANNON-WIENER DIVERSITY INDEX (H’)

The Shannon-Wiener Diversity Index (H’) (Equation 1) seeks to measure the degree of uncertainty in correctly predicting the species to which the next individual collected in successive sampling belongs (Martins and Santos, 1999). Considers equal weight between rare and abundant species (Magurran, 1988). The higher the value of H', the greater the diversity of the area under study.

$$H' = \frac{[N.\ln(N) - \sum_{i=1}^{S} n_i \ln(n_i)]}{N}$$

Where:
H' = Shannon-Wiener diversity index;
N = Total number of individuals sampled;
ni = Number of individuals sampled from the it species;
S = Number of species sampled;
ln = Neperian base logarithm (e).

The Shannon index is considered the most complete diversity index (Santos, 2009), as it gives the same weight to rare species as species with greater abundance. In general, the index value does not exceed 5.0, and when it does, only in situations where the number of individuals in the sample is greater than 105 (Magurran, 2004).

4.2.3 SIMPSON DOMINANCE INDEX (C)

Simpson's dominance index (Equation 2) measures the probability of two (2) individuals, selected at random from the sample, belonging to the same species (Brower and Zar, 1984). The estimated value of C varies from 0 (zero) to 1 (one), and for values close to one, diversity is considered greater.

$$l = \frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N-1)} : C = 1 - l$$

Where:
l = Dominance measure;
C = Simpson Dominance Index;
$n_i$ = number of individuals sampled from the it species;
N = total number of individuals sampled;
S = total number of species sampled.

## 4.3 METHODS FOR ESTIMATING FOREST DIVERSITY:

Most studies in the context of forest species diversity use field surveys via forest inventory and subsequently diversity indices to infer the diversity of a location/region. For example, Loschi et al. (2013) conducted a study in a fragment located in the Municipality of Itumirim, Minas Gerais, which presents a continuum of Mata de Galeria/Cerrado stricto sensu and evaluated species richness and diversity via sampling by the Shannon index. Lopes et al. (2023) also carried out data collection via sampling in three areas of phytophysiognomy of Cerrado sensu stricto, located in the northern region of the Legal Amazon, south of the state of Tocantins with the objective of comparing and evaluating alpha ecology using the indices of Shannon-Wienner diversity (H') and Pielou evenness (J'). Although traditional methods such as forest inventory are extremely important for conducting studies and collecting data, methods for estimating forest diversity have advanced significantly, incorporating innovations and approaches that integrate Remote Sensing (RS) data with field techniques (Fonseca et al., 2019). The interface between RS and fieldwork is crucial to validating and improving these methods.

## 4.4 REMOTE SENSING

In the context of remote sensing, forest diversity estimates can benefit from approaches that use passive and/or active remote sensors. Passive remote sensors record electromagnetic radiation reflected or emitted by the Earth's surface, while active sensors emit electromagnetic radiation directly to the surface and measure the return (Centeno, 2009).

### 4.4.1 LIDAR (LIGHT DETECTION AND RANGING)

The term LiDAR (Light Detection and Ranging) is used to designate a remote sensing technology that uses a laser pulse to characterize targets of interest in three dimensions (Wever and Lindenberger, 1999; Giongo et al., 2010). LiDAR combines laser light measurements and positioning to perform highly accurate surveys and mapping of surface objects and create a three-dimensional (3D) point cloud (NOAA, 2012).

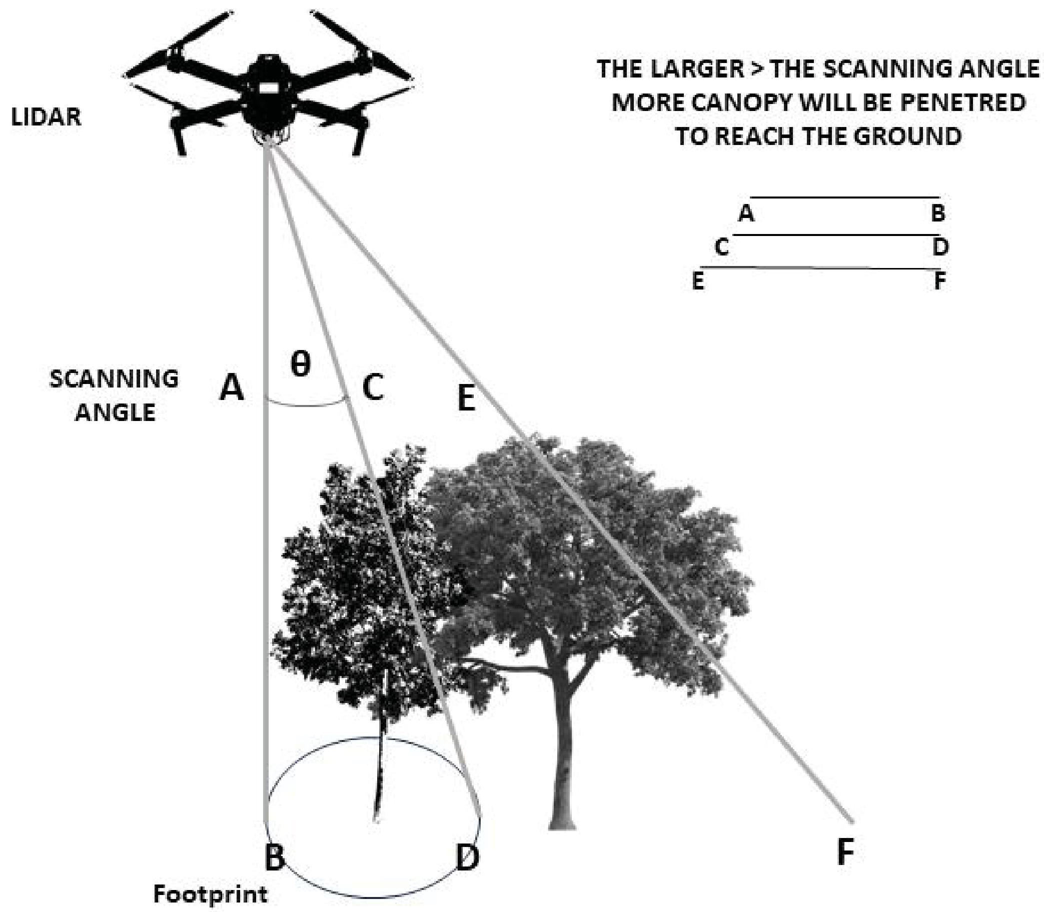
Laser scanning systems can be classified into two types:

i. static systems, which refer to the stationary condition in which the equipment collects data, remaining in a fixed position on the ground;
ii. dynamic systems, in which the equipment moves while obtaining data, and can be connected to a moving platform (Ferraz et al., 2016). Furthermore, systems with LiDAR sensors consist of a controller and a transmitter coupled to a platform (RPA, manned aircraft, satellite, etc.) in which laser light pulses are directed perpendicular to the line of flight by a mirror. scanning (Jensen, 2011). The pulses are emitted at high frequency, the laser's travel time is measured, from the transmitter to the target and back to the receiver, allowing the distance (range) to be calculated and a collection of three-dimensional points (x, y and z) to be obtained (Boland et al., 2004).

Systems with LiDAR sensors can be divided by the size of the footprint (area sampled by the laser pulse), the way they record the reflected signal and the sampling frequency and scanning pattern (Dubayah and Drake, 2000). The footprint is circular and depends on the topography and scanning angle (Baltsavias, 1999). For a canopy with uniform height and density, a greater scanning angle must be foreseen so that more vegetation is penetrated until the pulse reaches the ground (FIGURE 4) (Jensen, 2011).

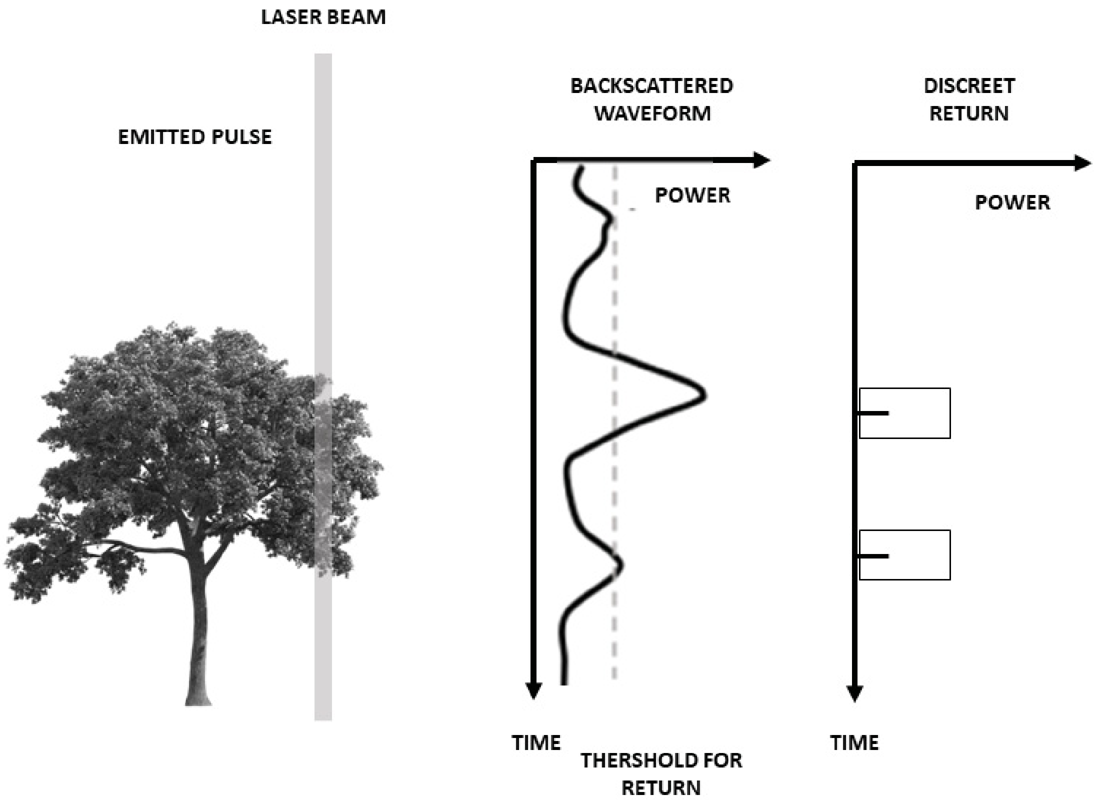
Systems with a small footprint may not be ideal for mapping forest structure, as they may excessively sample lower parts of the canopy, missing the top and soil, and requiring large areas to be flown extensively (Dubayah and Drake, 2000). On the other hand, the same authors stated that systems with a footprint of approximately an average crown diameter (10 – 25 meters) enable energy to reach the ground even in dense forests, allowing the tops of trees to be mapped, providing broad mapping areas.

FIGURE 4 - IMPACT OF A HIGH SCANNING ANGLE. SOURCE: ADAPTED FROM JENSEN (2011).




LiDAR sensors can make measurements of one or multiple returns for each signal peak per emitted pulse, either called discrete returns or providing full-wave representation of a return signal at fixed time intervals (Dong and Chen, 2018) (FIGURE 5). Full waveform systems are more expensive and describe more detailed information about the physical structure of the target (Kershaw et al., 2017), and are mainly used in forestry applications (Dong and Chen, 2018).

FIGURE 5 - PRINCIPLES OF A FULL WAVE (A) AND DISCRETE RETURN (B) SYSTEM. SOURCE: ADAPTED FROM CHAZETTE ET AL. (2016).




Because it can obtain multiple returns from different targets (leaves, branches, stem, soil surface), LiDAR technology is useful for silvicultural applications, allowing the derivation of tree height, canopy coverage and vertical structure of the stand (Köhl et al. al., 2006). Furthermore, such systems, as they can extract the vertical structure of forest formation, allow estimates of carbon stock in ecosystems with moderate to high biomass content, unlike optical or radar sensors, which tend to present saturation problems (Jensen, 2011).

According to Dubayah et al. (2020) LiDAR remote sensing is a very accurate way of investigating the vertical structure of vegetation, leading to efforts being directed in this direction over recent years. Some NASA missions can be highlighted, such as the Shuttle Laser Altimeter (SLA), ICESat, ICESat2 and Global Ecosystem Dynamics Investigation (GEDI) (Dubayah et al., 2020).

### 4.4.2 GEDI (GLOBAL ECOSYSTEM DYNAMICS INVESTIGATION)

GEDI (Global Ecosystem Dynamics Investigation) is an example of an instrument with full waveform space-based LiDAR technology on board the International Space Station (ISS) that collects samples of vegetation structure with a spatial resolution of approximately 25 m. The instrument consists of three lasers that produce eight parallel observation tracks, with each laser firing 242 times per second, illuminating a region of 25 m in diameter (footprint) and therefore obtaining the 3D structure of the site (Dubayah et al., 2020).

Each GEDI footprint is separated by 60 m along the trail, with a transverse distance of approximately 600 m between each of the eight trails (GEDI ECOSYSTEM LIDAR, 2018). GEDI lasers operate at a wavelength of 1064 nm (near infrared), and the pulses have a power of 10 mJ, being fired at an interval of 14 ns. The GPS allows you to locate GEDI's orbit in relation to the Earth's surface, while the star tracker components provide information about the instrument's orientation (GEDI ECOSYSTEM LIDAR, 2018).

GEDI data is available at different levels of processing: from the raw waveform, canopy height, vertical profile, canopy coverage, terrain altitude, among others (Dubayah et al., 2020). TABLE 1 shows the four product options available at different processing levels.

TABLE 1 - GEDI MISSION PRODUCTS AVAILABLE AT DIFFERENT PROCESSING LEVELS. SOURCE: ADAPTED FROM DUBAYAH ET AL., (2020).

| **Processing level** | **Product** |
|---|---|
| 1A | *Raw Waveform* |
| 1B | *Geolocated Waveform* |
| 2A | Terrain elevation, canopy top height, relative height metrics |
| 2B | Crown coverage fraction, crown coverage profile, leaf area index |

According to Healey et al. (2020) data collected by orbital LiDAR has the potential to provide canopy height measurements satisfactory enough to train regionally different height models. Fayad et al., (2021) used different metrics from GEDI observations to estimate dominant height (HD) and volume at plot level, in eucalyptus plantations in Brazil. Results showed that, due to the high accuracy of HD and volume estimation, GEDI is an excellent source of information to calibrate, validate and supplement future radar missions such as the BIOMASS satellite aiming to obtain continuous and high-resolution maps of ecosystem properties. forestry.

4.4.2 SENTINEL-2

The Sentinel-2 mission, a space component of the Copernicus family, comprises two twin polar orbiting satellites (Sentinel-2A and Sentinel-2B) placed in the same sun-synchronous orbit, equipped with a multispectral imager (MSI) that aims to monitor changes on the Earth's surface. The Sentinel-2A and Sentinel-2B satellites were launched in June 2015 and March 2017, respectively.

The units of this mission are at an altitude of 786 km, have a sun-synchronous orbital plane, a phase difference of 180º between them, a width of the imaging swath of 290 km and a revisit time of five days at the equator, considering both satellites (ESA, 2020). The limits of satellite coverage are between latitudes 56º S and 84º N and the satellites cross the equator at 10:30 am local time (Drusch et al., 2012).

Each of the mission units carries an MSI sensor with 13 spectral bands, four bands with a spatial resolution of 10 m, six bands with 20 m and three bands with 60 m (TABLE 2) (Drusch et al., 2012; ESA, 2020). The images are made available free of charge, orthorectified and 26 converted into reflectance at the top of the atmosphere (Level-1C product) (ESA, 2020).

The Sentinel-2A mission includes the MSI electro-optical sensor, with 13 spectral bands ranging between 400 and 2200 nm. The spatial resolution is 10 meters for four bands, 20 meters for six bands and 60 meters for three bands (ESA, 2021d). The Sentinel-2A constellation allows a 5-day revisit period (ESA, 2021e). The availability of bands in the visible, red edge and infrared ranges allows different vegetation indices to be explored. Level-2A products are delivered in surface reflectance and orthorectified images, available on the GEE platform.

TABLE 2 – CHARACTERISTICS OF THE SENTINEL-2A SPECTRAL BANDS.

| Band Number | Band's name | Central Wavelength (nm) | Spatial Resolution (m) |
|---|---|---|---|
| 1 | Aerossol | 443 | 60 |
| 2 | Blue | 490 | 10 |
| 3 | Green | 560 | 10 |
| 4 | Red | 665 | 10 |
| 5 | Red Edge 1 | 705 | 20 |
| 6 | Red Edge | 740 | 20 |
| 7 | Red Edge | 783 | 20 |
| 8 | NIR* | 842 | 10 |
| 8a | Red Edge | 865 | 20 |
| 9 | Water Vapor | 940 | 60 |
| 10 | Cirrus | 1.375 | 60 |
| 11 | SWIR 1 | 1.610 | 20 |
| 12 | SWIR 2 | 2.190 | 20 |

* near-infrared

### 4.4.3 ERA5 RE-ANALYSIS

The “ECMWF Re-Analysis” (ERA) project developed by the European Center for Medium-Range Weather Forecasts (ECMWF) brings together reanalysis data from a large number of climatological observations worldwide. Meteorological reanalysis can be defined as a gridded dataset. These are obtained through the combination of data acquisitions measured by meteorological institutes and physical models of global circulation and forecasting. The result of this interaction is a new, high-resolution dataset that can describe in detail the climate behavior of different meteorological variables, atmosphere, land surface and oceans (ECMWF, 2020).

ERA5, for example, is part of this project. The record of data assimilated by ERA5 from 1950 to the present day. The information made available to users are hourly estimates of many climatological variables acting on atmospheric, terrestrial and oceanic conditions.

ERA5 presents several improvements when compared to its predecessors, ERA-Interim, including horizontal spatial resolution of 31 km, with data gathered in a grid divided into 137 levels, from the earth's surface up to 80 km height of the atmospheric layer. The data is updated hourly and includes error estimates. It also presents curves on the evolution of greenhouse gas emissions, surface temperatures of oceans and glaciers and volcanic eruptions (ECMWF, 2020).

#### 4.4.4 SHUTTLE RADAR TOPOGRAPHY MISSION (SRTM)

The Shuttle Radar Topography Mission (SRTM) is the result of a collaborative effort between NASA and the National Geospatial-Intelligence Agency (NGA) and the participation of both German and Italian space agencies. The purpose of the SRTM was to generate, at the beginning of the 21st century, a digital elevation model (DEM) of the Earth using Synthetic Aperture radar (SAR) interferometry, with a one arc-second grid spacing (~30m) and approximately 15 m accuracy (Farr et al., 2007).

The SRTM provides detailed altimetry data, generating DEM that can help characterize the terrain and identify areas suitable for different vegetation types.

### 4.5 MAIN RESEARCH FOR ESTIMATION OF DIVERSITY INDEXES BY REMOTE SENSING IN THE CERRADO

Combining data from different sensors and techniques (optical, radar, LIDAR) is becoming more common to obtain more complete information about forest diversity (Cabacinha and De Castro, 2009). Using machine learning algorithms to process large RS datasets is enables more accurate identification of species, vegetation structure, and changes over time (Carneiro, 2023). The convergence between advanced RS methods and fieldwork is boosting the accuracy of forest diversity estimates, enabling a deeper understanding of ecosystems and their conservation status.

Cabacinha and De Castro (2009) analyzed correlations between the floristic diversity of forest fragments, based on Shannon's diversity indices and Pielou's evenness, with patch metrics and vegetation indices derived from Landsat OLI sensor images acquired in two seasons. The authors' findings revealed that the index that showed the highest correlation with floristic diversity was the EVI and the one that showed the highest correlation with evenness was the MVI5. Furthermore, the diversity and evenness indices showed correlations with the fragment shape when analyzed in relation to patch metrics.

Radar data were also evaluated to characterize the Shannon-Wiener diversity index in the Cerrado. Gomes and Maillard (2006) found positive correlations in the dry season for the Shannon index (r = 0.384) representing floristic diversity. Batista et al., (2016) using ordinary kriging analyzed the spatial distribution and behavior of species richness and diversity in a fragment of Cerrado field, in 2003 and 2014. The results showed strong spatial dependence between species richness and diversity. species by the Shannon-Wiener diversity index (<25% degree of spatial dependence). In the study, it was possible to identify areas of low and high diversity and species richness in the Campo Cerrado fragment.

More recently, Machado et al. (2019) tested whether spectral variables from the Landsat imagery could be used as indicators of plant species diversity in the Caatinga. The authors found that there is a positive correlation between richness and the near-infrared (NIR) spectral band. Furthermore, it was demonstrated in the study that the NIR spectral band was also responsible for better explaining the variation in leaf level reflectance among eight species that occur in the region. Therefore, the NIR band variable can be used as an indicator of species richness.

Righi et al. (2023) verified the relationship between the biodiversity of the flora existing in the Cerrado Stricto Sensu with its aerial biomass and carbon stocks. The results showed that total aerial biomass presented a highly significant asymptotic relationship with biodiversity, demonstrating its importance in achieving high biomass accumulation. Mapfumo et al. (2016) tested the relationship between species diversity measured in situ with the Normalized Difference Vegetation Index (NDVI) and the NDVI Coefficient of Variation (CVNDVI) derived from high and medium spatial resolution satellite data space in dry, humid and coastal savannas. The results of the study suggested that tree species diversity can be successfully predicted using satellite-derived vegetation indices such as NDVI and CVNDVI.

Mutowo and Murwira (2012) tested how tree species diversity at three forest savanna sites is related to indices derived from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) satellite images. The authors tested the use of standard deviation of near-infrared radiance (NIR stdev) and tree canopy cover estimated using the soil-adjusted vegetation index (SAVI) in estimating diversity. The results show that tree species diversity has a significant ($p<0.05$) bump-shaped response to variations in standard NIR radiance and SAVI. Furthermore, the results show that the combination of stdev NIR and SAVI explains between 60% and 64% of the variations in tree species diversity, an improvement between 30% and 54% explained by the indices individually. The authors concluded that remote sensing data from ASTER can be successfully used to estimate tree species diversity in savannah forests.

Madonsela et al. (2017) explored the utility of spectral information across the Landsat-8 spectrum using Principal Component Analysis (PCA) to estimate alpha diversity (Shannon (H′), Simpson (D2) and richness of species (S)) in the savanna forest of southern Africa. The results of the study indicate that: (i) the measurement scales of vegetation indices impact their sensitivity to vegetation characteristics and their ability to explain the diversity of tree species; (ii) principal components increase the utility of Landsat-8 spectral data for estimating tree species diversity and (iii) species diversity indices that consider species richness and abundance (H′ and D2) relate to better with the Landsat-8 spectral variables.

Gyamfi-Ampadu et al. (2021) evaluated the influence of spectral and spatial resolutions of PlanetScope, RapidEye, Sentinel-2A and Landsat 8 images on diversity prediction based on the Shannon diversity index (H′), the diversity index of Simpson and species richness. The results found showed that Sentinel-2 was the best image, producing the highest coefficient of determination ($R^2$) in both the Shannon Index ($R^2 = 0.926$) and species richness ($R^2 = 0.923$). Both Sentinel-2 and RapidEye produced comparable higher accuracy for the Simpson Index ($R^2 = 0.917$ and $R^2 = 0.915$, respectively). PlanetScope was the second most accurate for species richness ($R^2= 0.90$), while Landsat 8 was the least accurate for the three diversity indices. The results of this study suggest that both spectral and spatial resolution influence the forecast accuracy of satellite images.

## 5 MATERIALS AND METHODS

### 5.1 STUDY AREA

The study area spans the states of Minas Gerais and Goiás, covering the savanna forest strips in Cerrado, Brazil (Fig. 6). The area is divided into two land management regimes, that is, communal areas (Federal University of São João del-Rei - USJF) and protected areas (Serra do Cipó National Park - SCNP, Chapada dos

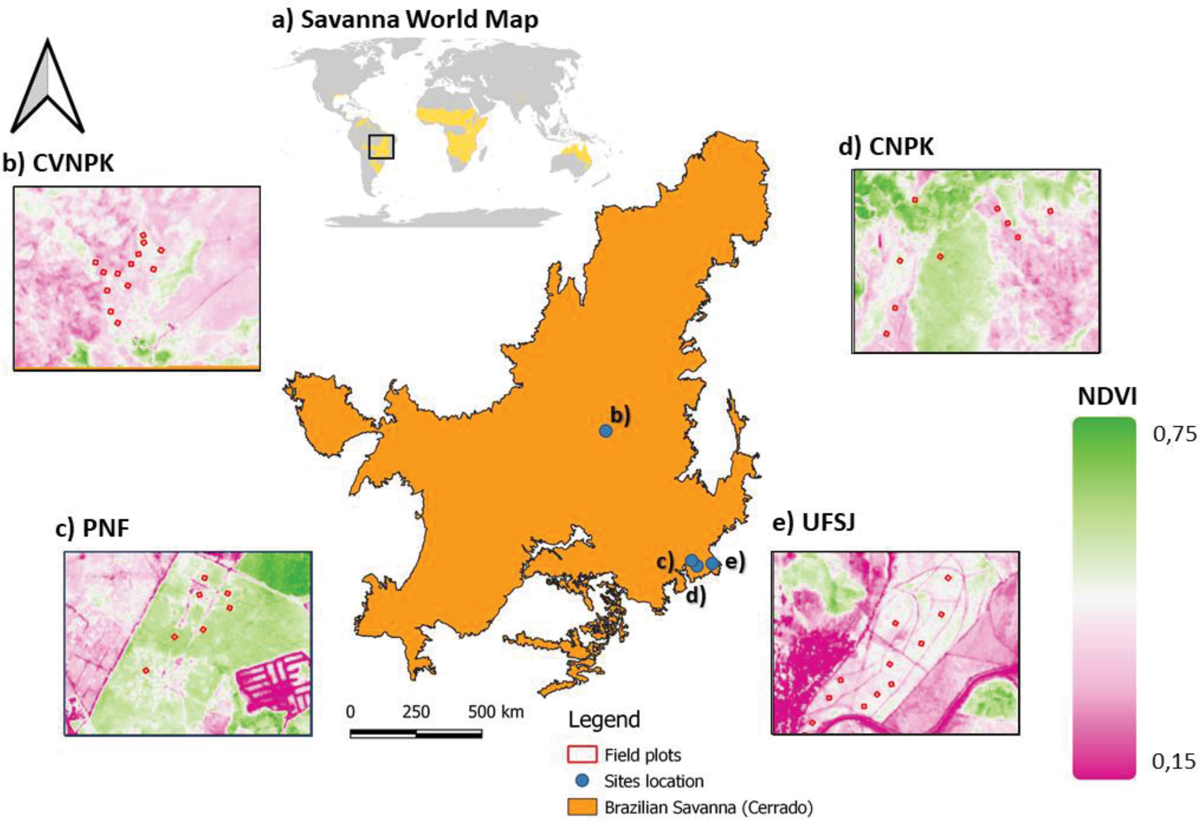
Veadeiros National Park - CVNP, Paraopeba National Forest - PNF) with different land-use practices.

The PNF, UFSJ, and SCNPK study sites are in the southeast portion of the Cerrado, state of Minas Gerais, while the CVNPK is located in the central portion of the Cerrado (FIGURE 6). Each site is characterized by different climatic, topographic, and water regime conditions, in addition to contemplating different vegetation formations present in the Cerrado. More detailed information can be found in Costa et al. (2021). Overall, open grasslands are characterized by the predominance of herbaceous species and some shrubs, with the absence of trees. Savanna formations refer to areas with trees and shrubs scattered over a stratum of grasses and herbs, with the absence of continuous tree canopies. Forest formations are areas with a predominance of tree species and may have continuous or discontinuous canopy (Ribeiro and Walter, 2008).

More specifically, in the UFSJ forest (19°28'S, 44°11'W) the predominant vegetation is Cerrado sensu stricto characterized by the dominance of trees with scattered shrubs and grass understorey. The PNF (19°20'S and 44°20'W) is comprised of 150 ha remnants of Cerrado vegetation, including both savanna (a.k.a., "Cerrado sensu stricto") and forest formations (a.k.a., "Cerradão") (Neri et al., 2013). The landscape in the CVNPK (13°51'-14°10'S, 47°25'-42'W) is formed by mosaics of different vegetation types (Ribeiro and Walter, 2008) characterized by a predominance of forest formations at low elevations and Cerrado with montane savannas at high elevations (Felfili et al., 2007). Wet and dry grasslands and savannas occur between streams, covering most of the landscape. At the northwest edge of the park, dry deciduous forests are found, whereas at the southwest edge riparian evergreen forests are most common (Flores et al., 2020). In total, the CVNPK represents 77% of the savanna formation, and about 10% corresponds to the forest fragments (Porto et al., 2011). Finally, the vegetation in SCNPK varies and comprises different physiognomies, from open grasslands ("Campo Limpo") at altitudes below 1,000 m to savanna formations with different proportions of woody cover ("Campo Sujo", "Campo Cerrado" and "Cerrado sensu stricto") and forest formations ("Cerradão"), all classified as part of the Cerrado sensu lato (Oliveira-Filho and Ratter, 2002). Additionally, the rupestrian grasslands are found at elevations above 1000m (Benites et al., 2003).

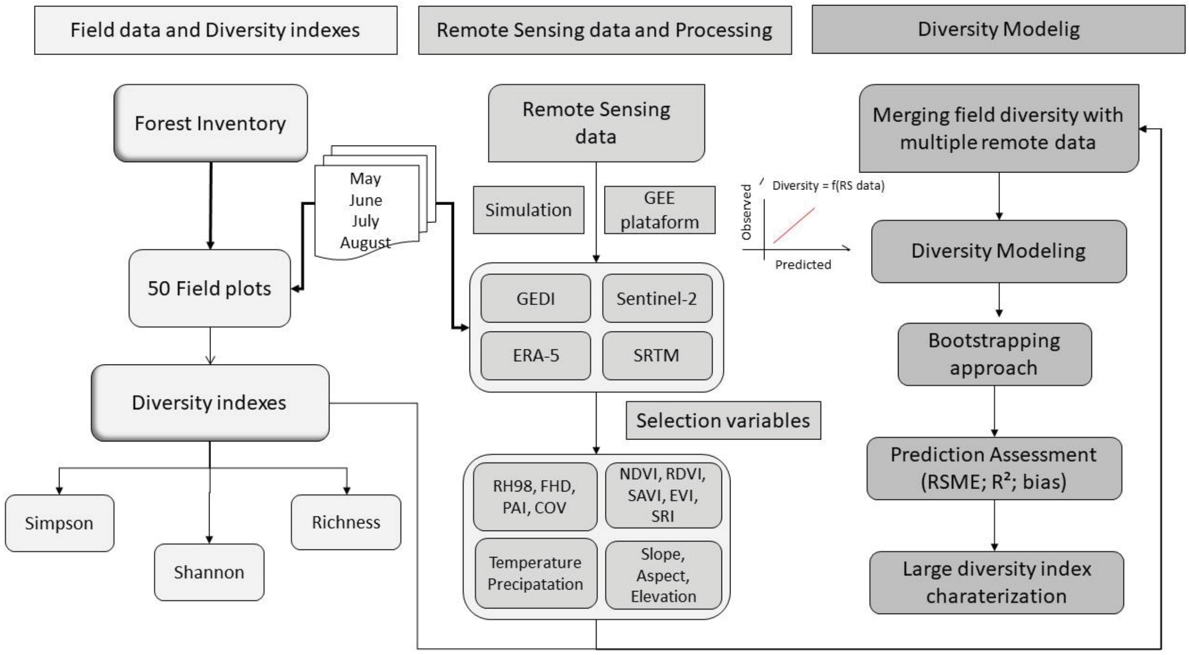
For the characterization of the vegetation types in the Cerrado, herein we follow the definitions proposed by Ribeiro and Walter (2008), which subdivide Cerrado into open grasslands ("Campo Sujo", rupestrian grasslands and "Campo Limpo"), savanna formations (campo Cerrado, Cerrado sensu stricto, palm grove and veredas (wet savannas) and forest formations (riparian forest, gallery forest, dry forest and cerradão). In addition, the Cerrado physiognomic gradient is related to environmental factors and is maintained by spatially and temporally dynamic disturbance regimes, both natural and anthropogenic (Durigan et al., 2018).

FIGURE 6 - SPATIAL LOCATION OF THE BRAZILIAN SAVANNA (CERRADO) AND STUDY SITES WHERE UAV-LIDAR AND FIELD DATA WERE COLLECTED, NAMELY. D) SERRA DO CIPÓ NATIONAL PARK (SCNPK), B) CHAPADA DOS VEADEIROS NATIONAL PARK (CVNPK), C) PARAOPEBA NATIONAL FOREST (PNF), AND




An overview of the methodology is illustrated in FIGURE 7. The approach was divided into three phases: (1) Field surveys and calculating diversity indices, (2) remote sensing and processing data collection, and (3) development of predictive models and validation.

FIGURE 7 - METHODOLOGICAL FLOWCHART FOR PROCESSING MULTISOURCE REMOTE SENSING DATA AND MODELING DIVERSITY IN DIFFERENT LANDSCAPES IN THE CERRADO.




## 5.2 FIELD AND FOREST DIVERSITY DATA

In total, 50 square plots, 900 m² (30×30 m) in area, were measured between June and July of 2019 for this study. Plot corners were registered using a Differential Global Navigation Satellite System (DGNSS). Each tree was taxonomically identified, and its height (ht, in m) and diameter at breast height (dbh, in cm) were measured using a clinometer and diameter tape, respectively. After fieldwork, the diversity indices (Species richness, Simpson index, and Shannon–Wiener index) were calculated for each plot (TABLE 3).

TABLE 3 - DIVERSITY INDICES USED IN THE STUDY AND THEIR EQUATIONS, WHERE PI IS THE PERCENT COVER PROPORTION OF THE SPECIES.

| Diversity metric | Equation | References |
|---|---|---|
| Species richness (S) | Number of species | Colwell (2009) |
| Shannon's index (H′) | $H' = \sum pi.LN\ pi$ | Shannon and Weaver (1963) |
| Simpson's index (D) | $D = \sum pi^2$ | Simpson (1949) |

Species richness refers to the total number of different species in a sampling unit (Oldeland et al., 2010). Other diversity indices combine two attributes of a community: species richness and equability (Hurlbert 1971; Peet, 1974). Equability refers to how similarly species are represented in the community. If all species have the same representativeness (or importance; Peet, 1974), the equability will be maximum. Most diversity indices are said to be non-parametric as they do not depend on the parameters of a distribution. They usually consist of simple mathematical expressions involving the relative abundance of each species in the sample (Melo, 2008).

The Shannon-Wiener diversity index, represented by H', is calculated based on the number of individuals in each species and the total number of individuals sampled. As one of the most frequently used diversity indices, it is sensitive to both species rarity and abundance, and has been used in different studies as a measure of alpha diversity (Madonsela et al., 2017; Oldeland et al., 2010). The Simpson index is also a widely used measure of species diversity (Nagendra, 2002; Lamb et al., 2009) that considers the abundance and number of species present in an area and estimates the probability that two individuals chosen at random belong to the same species.

## 5.3 REMOTE SENSING DATA

In this study, four data sources including Sentinel-2 optical imagery, ERA-5 climate data, SRTM-DEM imagery, and GEDI simulated data were selected for the characterization of diversity in our four studies areas. For the first three data sources, we used the Google Earth Engine (GEE) platform for data collecting and processing, while the GEDI data were simulated using the rGEDI package in R (R Development Core Team, 2020).

## 5.4 UAV-LiDAR GATOREYE

We simulated GEDI data from the UAV-lidar 3D point cloud for calibrating species diversity models to avoid the geolocation errors of GEDI (~10–20 m) and due to the fact that GEDI orbits likely did not overlay our field plots. For that, the GatorEye UAV-LiDAR system (Broadbent et al., 2021) was used to scan our study sites for two weeks in July 2019, almost simultaneously with the field data collection. The GatorEye used a DJI M600 Pro flight platform mounted with a Phoenix Scout Ultra core to integrate LiDAR (VLP16) with an inertial motion unit (Novatel STIM 300), and centimeter accuracy differential GNSS system. A complete description of the GatorEye system can be found in a recent study (da Costa et al., 2021), and data are available on the GatorEye website [www.gatoreye.org]. For further information the reader is referred to Broadbent et al. (2021) and d'Oliveira et al. (2020). The autonomous flight was programmed to survey at a mean speed of 14 $m.s^{-1}$ at 100 m above ground level (a.g.l.), with flight lines spaced 100 m apart. In total, across the four study sites, we flew more than 600 km of flight lines, with a lidar swatch coverage of 1,854 hectares, which to our knowledge, as of the flight date, was the largest area of UAV-lidar used in a publication and the only one in this ecosystem. The final merged point clouds were about 100 GB in total size and had a very high-density of approximately 450 points $m^{-2}$ across all study sites (da Costa et al., 2021). Herein, UAV-LiDAR 3D point cloud data processing included implementing the GatorEye Multi-scalar Post-Processing Workflow (as detailed in Broadbent et al., 2021), aligning the flight lines, and clipping the point clouds within the field plots for GEDI data simulation (Leite et al., 2021).

## 5.5 NASA'S GEDI

GEDI data from the UAV-LiDAR 3D point cloud were simulated following methods of Leite et al., (2021). The GEDI pre-launch plan included the development of a GEDI simulator that can reproduce the on-orbit GEDI data characteristics for the calibration of aboveground biomass models (Hancock et al., 2019). The simulation includes transforming discrete-return lidar point clouds into full-waveform signals

(Blair and Hofton, 1999) in GEDI-sized footprints and adding the expected GEDI instrument noise. The waveform signal-to-noise ratio (SNR) on the on-orbit GEDI data depends on characteristics such as laser type (power or coverage), acquisition time (day or night), canopy cover and atmospheric conditions (Hancock et al., 2019; Dubayah et al., 2020a; Duncanson et al., 2020). The simulator ensures consistency across point cloud flight characteristics especially for high-density LiDAR point clouds, such as those used as input in this study. It allows consistently transferring models to the on-orbit GEDI data. A complete description and validation of the GEDI simulator are described in detail by Hancock et al. (2019).

In our study, GEDI-like full waveform data were simulated using the high-density UAV-LiDAR point clouds clipped to the study sample plots using the gediWFSimulator tool in the rGEDI package (Silva et al., 2020) in R (R Core Team 2020). We added realistic noise considering a beam sensitivity of 0.98 (i.e., the canopy cover at which ground is detected 90% of the time with a 5% probability of a false positive, see Hancock et al., (2019) by using a link margin of 4.956 at 95% of canopy cover that relates to noise of the power beam collecting data at night (Boucher et al., 2020). For ground detection and metrics calculation, the waveforms were denoised and smoothed by setting the noise threshold as the mean plus three standards deviations and smoothing width (applied after denoising) equal to 0.5 m (Qi et al., 2019, Silva et al., 2021). After simulating the GEDI waveforms, a suite of canopy metrics was computed based on the cumulative waveform energy (i.e., 10%, 25%, 50%, 75%, 98%, and 100%; Relative Height 10% (RH10), RH25, R50, RH75, RH98, RH100; COV: canopy cover (%); FHD – foliage height diversity). These metrics were calculated using the gediWFMetrics function in rGEDI (Silva et al., 2020) and selected to match the GEDI Level 2A and 2B products and facilitate model interpretability.

## 5.6 SENTINEL 2 MSI - LEVEL 2A

Sentinel-2 provides multi-spectral data, including four bands with 10 m spatial resolution, six bands with 20 m spatial resolution, three bands with 60 m spatial resolution, and three quality assessment (QA) bands, where QA60 is a bitmask frequency band with cloud mask information (Ogilvie et al., 2020). We employed the level 2A surface reflectance products of Sentinel-2, which are available in GEE (Dataset ID: ee.ImageCollection("COPERNICUS/S2_SR"). Images from May 1, 2019 to August 31, 2019 were selected with a cloud cover of less than 30% based on the "CLOUDY_PIXEL_PERCENTAGE" attribute, and we combined these images to minimize cloud impact. Herein, we computed the following vegetation indices (VIs): Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Soil Adjusted Vegetation Index (SAVI), Renormalized Difference Vegetation Index (RDVI) and Simple Ratio Index (SRI) (TABLE 4).

On GEE, we applied a compound median reducing function to calculate the median value of each image collection from May to August, i.e., Cerrado's dry season. The median reducing function removes clouds, which have high values, and shadows, which have low values, from the image. The output composite value is the median in each band over time. Then, we applied a clip function to group the image collections in the study region and then calculated the indices using equations.

TABLE 4 - VEGETATION INDICES DERIVED FROM IMAGES FROM THE SENTINEL-2 SATELLITE MSI SENSOR.

| Indices | Equations | Reference |
|---|---|---|
| Normalized Difference Vegetation Index | $NDVI= \frac{NIR\text{-}RED}{NIR+RED}$ | Rouse et al., (1973) |
| Renormalized Difference Vegetation Index | $RDVI= \frac{NIR\text{-}RED}{\sqrt{NIR+RED}}$ | Richardson and Wiegand (1977) |
| Soil Adjusted Vegetation Index | $SAVI= \frac{NIR\text{-}RED}{NIR+RED + 0.5} * 1 + 0.5$ | Huete (1988) |
| Enhanced Vegetation Index | $EVI = 2.5 * \frac{NIR - RED}{NIR + 6 * RED - 7.5 * A + 1}$ | Liu and Huete, 1995 |
| Simple Ratio | $SR = \frac{NIR}{RED}$ | Jordan (1969) |

The NDVI is one of the most commonly used remotely sensed spectral vegetation indices and is calculated from reflectance in the near infrared and red portions of the electromagnetic spectrum (Rouse et al., 1974). The Enhanced Vegetation Index (EVI) was proposed by Liu and Huete (Liu and Huete, 1995) to compensate for the limitations of the NDVI regarding soil background and atmospheric interference. In a general way, the NDVI is sensitive to the content of chlorophyll and other pigments responsible for absorbing solar radiation in the red range of the electromagnetic spectrum, while the EVI is also more sensitive to the variation in canopy structure, including the Leaf Area Index (LAI), the plant physiognomy, and the canopy volume because of the inclusion of blue band information (Gao et al., 2000; Huete et al., 1988).

The RDVI was proposed to combine the advantages of the Difference Vegetation Index (DVI=NIR−Red; Jordan, 1969) and the NDVI for low and high LAI values, respectively. Besides, the RDVI was proposed to minimize the saturation effect. The Soil-Adjusted Vegetation Index (SAVI; Huete, 1988), was proposed to account for changes in soil optical properties. The SAVI includes a canopy background adjustment factor L. Finally, the Simple Ratio (SR) index was derived as one of the commonly used vegetation indices (Tucker, 1979; Sellers, 1985). It provides unique information, which is not available in any single band. It is used for discriminating between soil and vegetation in the study region (Erener, 2011).

## 5.7 ANCILLARY DATA

We used ERA-5 (Hoffmann et al., 2019) and the Digital Elevation Model (DEM) obtained from the Shuttle Radar Topography Mission (SRTM) data (Mukul et al., 2017; Yang et al., 2011) as ancillary data for forest diversity modeling. The ERA-5 (fifth generation) is a latest climate reanalysis produced by ECMWF (European Centre for Medium-Range Weather Forecasts) / Copernicus Climate Change Service (Škerlak et al., 2014), with a spatial resolution of 31 km. This dataset is freely available and offers a detailed overview of the atmosphere. In addition, the ERA5 is part of GEE’s datasets consisting of air temperature as a monthly average at 2 m height with availability from 1979 to present. In this way, we selected the products "mean_2m_air_temperature" that corresponds to the average air temperature at 2m height (monthly average), as well as the "total_precipitation", which refers to the total precipitation (monthly sums). These were chosen as the variables to represent the temperature and precipitation of the study areas.

SRTM data, measured and released by NASA and the National Surveying and Mapping Bureau of the Department of Defense, cover 80% of the global land surface

(Farr et al., 2007). The SRTM Version 3 (V3) product, provided by NASA Jet Propulsion Laboratory (JPL) at a resolution of 1 arc second (approximately 30 m) was exploited in our research. With the code available on the GEE platform (https://developers.google.com/earth-engine/datasets/catalog/USGS_SRTMGL1_003), we used functions to directly extract the aspect, slope and elevation for our study areas.

TABLE 5 - SETS OF REMOTE SENSING CANDIDATE METRICS FOR THE FOREST SPECIES DIVERSITY MODELING.

| Metric set | Source name | Variables |
|---|---|---|
| 1 | GEDI | RH98 + FHD + PAI + COV |
| 2 | SENTINEL-2 | NDVI + SR + SAVI + RDVI + EVI |
| 3 | ERA-5 | Temperature + Precipitation |
| 4 | SRTM | Slope + Aspect + Elevation |

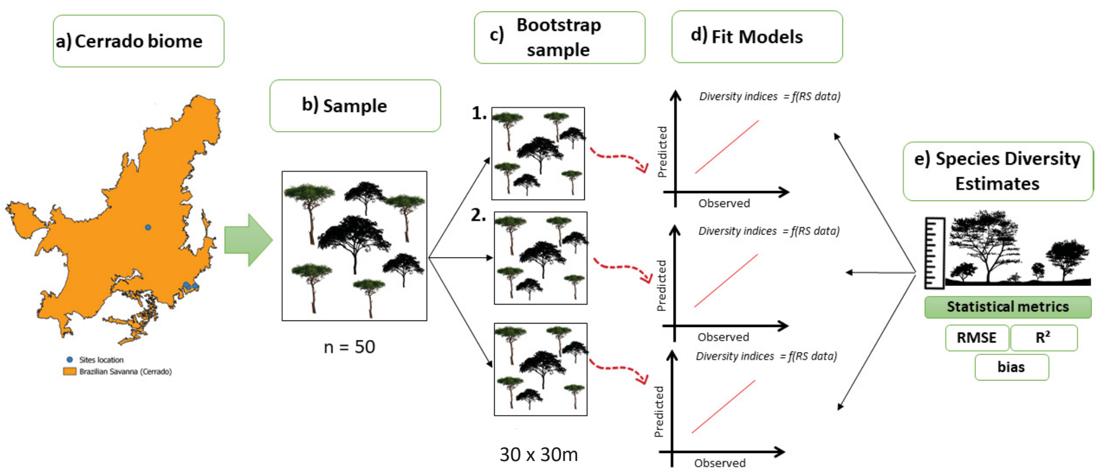
5.8 FEATURE SELECTION

Feature selection is one of the critical steps in machine learning and has two main functions: (1) reducing the number of features and dimensions, thereby enhancing model generalizability and reducing overfitting, and (2) enhancing the understanding between features and eigenvalues. The variable selection using the random forest (VSURF) algorithm was applied to the original dataset to determine the appropriate variable number. VSURF is a wrapper-based algorithm that uses random forests as the base classifier (Genuer et al., 2010). Feature variables are first ranked based on a variable importance measure, and low-scoring features are eliminated to reduce the number of features and improve model accuracy (Genuer et al., 2015). In the final step, a ranked list of only the most important features is produced.

Then, we applied Random Forest (RF) regression (Breiman, 2001) for predicting diversity indices. RF has been widely used in forest modeling based on earth observation data (Rex et al., 2020; Ghosh et al., 2018; Torbick et al., 2016) due to its non-parametric nature and ability to deal with dimensionality, multicollinearity and overfitting (Rodriguez-Galiano et al., 2012; Belgiu and Drăguţ, 2016). In RF modeling, two parameters are required to construct the decision trees. The first one is the number of decision trees that should be generated. The second is the number of variables that need to be selected for the greatest split when the trees become larger over a period (k predictor). The RF algorithm was implemented in R using the randomForest package (Liaw and Wiener, 2002).

The bootstrapping approach was applied to assess the models' precision and accuracy. First, we completed 1000 random permutations of the original data, and then split the data to a training set and a test data. 2/3 of the data were used to train the models, and the remainder was used to assess their predictive ability. The strength of the relationship was assessed using the coefficient of determination ($R^2$), and the Imodel's performance was assessed using the root mean square error (RMSE) and bias. The diversity models were based on multi-source data from remote sensing and in-situ measurements. All diversity measures and regressions were calculated using RStudio v1.4.17 (R Development Core Team, 2020). Figure 8 shows the bootstrapping process for modeling diversity indices in the Brazilian Cerrado.

FIGURE 8 - FLOWCHART PRESENTING THE BOOTSTRAPPING METHOD FOR MODELING DIVERSITY INDICES IN THE BRAZILIAN CERRADO; A) CERRADO AND FIELD PLOTS; B) AREA SAMPLING (TOTAL 50 PLOTS); C) FROM THE SAMPLE - BOOTSTRAP METHOD WITH A TOTAL OF 1000 REPETITIONS; D) FIT MODELS DEVELOPED AFTER BOOTSTRAP SAMPLING FOR EACH DIVERSITY INDEX; E) VALIDATION OF FITTED MODELS AND STATISTICAL METRICS FOR EACH DIVERSITY MODEL.



## 5.9 DIVERSITY INDICES CHARACTERIZATION IN THE CERRADO

To characterize the Cerrado species diversity indices, we carried out a series of sequential processes. First, the GEDI Level 2A and 2B version 2 data products (Dubayah et al., 2021a; 2021b) collected between April 18, 2019 and March 01, 2023 were downloaded over the entire Cerrado vegetated area. The GEDI orbits intersecting Cerrado limits were found and downloaded using the gedifinder and gediDownload functions in the rGEDI package (Silva et al., 2020).

Then, in GEE, we stack all raster layers (Sentinel-2, SRTM and ERA-5) using the layerstack function. With the stacked optical data, in the R environment, we imported the GEE data and added it to the GEDI data footprint. In this way, each GEDI footprint additionally received the values for each feature of the optical data. That is, in a given GEDI footprint, in addition to the previously selected structural metrics, we also have all Sentinel-2, ERA-5 and SRTM variables.

The footprints were masked to the Cerrado vegetated area based on the land cover classification from MapBiomas for the same year of the data collection (Souza et al., 2020). This procedure guarantees that only pixels of forest, savanna and grassland vegetation were mapped in the following steps. The GEDI footprint-level metrics (Table 2) were extracted using the getLevel2AM and getLevel2B functions and filtered using the quality flag (quality_flag = 1). This flag indicates usable data by summarizing individual quality assessment parameters based on waveform shot energy, sensitivity (< 0.9 over land), amplitude, and real-time surface tracking quality (Hofton and Blair, 2019; Beck et al., 2020).

Finally, with all the data stacked, filtered and adjusted to the established parameters the diversity indices models developed were applied to the GEDI footprints (diameter of ~25 m) which combines the GEDI and optical data collected across the Cerrado biome extent. Diversity indices maps of each component were created by taking the average of the footprint-level estimates at 1-km2 grid cells for mapping purposes and compatibility with planned gridded GEDI products (Dubayah et al., 2020) and requirements for global biomass maps (Hall et al., 2011). Finally, we calculated the uncertainty of diversity indices predictions in each cell by accounting for the footprints' variability within the cell, uncertainty associated with the RF algorithm, and RF lack of fit. To do this, we followed the same workflow developed in (Leite et al., 2021).

# 6 RESULTS

## 6.1 PREDICTIVE MODELS FOR SPECIES DIVERSITY INDICES

First, we evaluate the diversity index models for each data set separately. That is, diversity index models for the GEDI metrics set, diversity index models using the information from Sentinel-2, diversity index models for the SRTM dataset and finally, diversity index models for the information of ERA-5. The performance results of each model in terms of validation are shown in Table 6.

TABLE 6 - CROSS-VALIDATION PERFORMANCE ASSESSMENT IN 500 ITERATIONS OF MODELS USED TO ESTIMATE THE DIVERSITY INDEX (SHANNON, SIMPSON AND RICHNESS) USING RANDOM FOREST, GEDI WAVEFORM METRICS AND CONVENTIONAL PASSIVE OPTICAL IMAGING AS PREDICTORS. $R^2$ = COEFFICIENT OF DETERMINATION; RMSE = ROOT MEAN SQUARE ERROR; AND BIAS.

| Model | Diversity Index | $R^2$ | RMSE | RMSE% | Bias | Bias % |
|---|---|---|---|---|---|---|
| GEDI | Shannon | 0.30 | 0.59 | 31.63% | 0.2 | 10.00% |
| | Simpson | 0.34 | 0.20 | 26.30% | 0.09 | 12.00% |
| | Richness | 0.24 | 4.54 | 44.02% | 0.3 | 2.99% |
| SENTINEL-2 | Shannon | 0.52 | 0.6 | 37.00% | 0.2 | -12.00% |
| | Simpson | 0.15 | 0.18 | 27.50% | -0.05 | -8.59% |
| | Richness | 0.46 | 4.30 | 47.52% | -1.67 | -18.49% |
| SRTM | Shannon | 0.75 | 0.56 | 35.17% | -0.17 | -11.20% |
| | Simpson | 0.69 | 0.23 | 36.00% | -0.05 | -7.35% |
| | Richness | 0.86 | 3.22 | 33.41% | -0.92 | -10.00% |
| ERA-5 | Shannon | 0.24 | 0.87 | 61.65% | -0.25 | -18.21% |
| | Simpson | 0.20 | 0.34 | 58.83% | -0.1 | -18.75% |
| | Richness | 0.30 | 4.88 | 66.02% | -1.38 | -18.70% |

In general, the results of the models in terms of validation presented a low to moderate performance (0.15-0.52) in relation to $R^2$, except for the models that used topographic variables from the SRTM, for these models the performance in relation to $R^2$ ranged from 0.69-0.86.

The models that used GEDI structural metrics showed similar performance. For $R^2$, the models presented a performance that ranged from 0.24-0.34. The Simpson model was the model that generally presented the best performance with GEDI metrics ($R^2$ = 0.34, RMSE = 26.30% and Bias = 12%).

For models that used Sentinel-2 variables, the performance was opposite to models with GEDI variables. For these models, the Simpson index presented the lowest performance in terms of $R^2$ (0.15). However, Simpson's model presented the best results in relation to the RMSE and BIAS metrics (27% and -8.59%, respectively). The Shannon and Richness models showed moderate performance in relation to $R^2$ (0.52 and 0.46, respectively). The Shannon model presented values of RMSE = 37% and Bias % = -12%, while the Richness model presented an RMSE = 47.53% and Bias = 18.49%.

For the models that used SRTM variables, the performance of the models for all indices presented a good fit in terms of $R^2$. Simpson's model presented an $R^2$ = 0.69, Shannon = 0.75 and Richness = 0.86. Regarding the RMSE and Bias metrics, the performance was similar with the models that used GEDI and Sentinel-2 variables. Shannon (RMSE= 35.17% and Bias = -11.20%), Simpson (RMSE= 36% and Bias = -7.35%) and Richness (RMSE= 33.41% and Bias = 10.00%).

Finally, the models that considered the ERA-5 variables presented a lower performance compared to the other models from the different data sets. Overall, for $R^2$ the models presented a low performance (0.2-0.3). Furthermore, in relation to RMSE, the models varied between 58%-66%, and for Bias the performance of the indices was close to -18%.

## 6.2 SPECIES DIVERSITY INDICES AND REMOTE SENSING METRICS

Different types of species diversity indices showed similar patterns in VSURF variable selection method. The number of selected features differed between different types of indices (TABLE 7). For the Shannon index, the ideal number of variables was 5, while 4 variables were selected for the Simpson index and for species Richness (FIGURE 9).

TABLE 7 - VARIABLE SELECTION USING RANDOM FORESTS (VSURF) FOR EACH DIVERSITY INDEX MODEL.

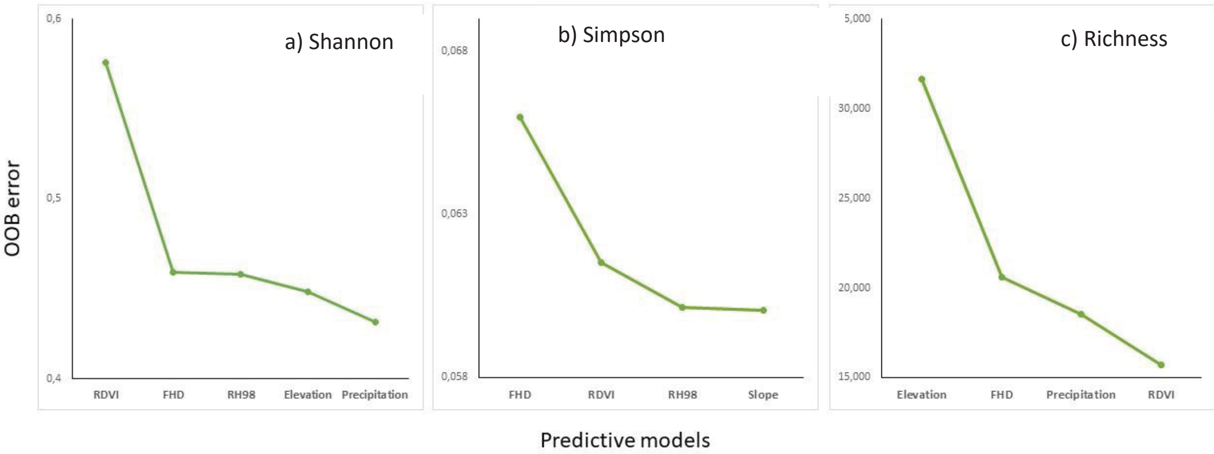
| Diversity index | Selected variables |
|---|---|
| Shannon | RDVI + FHD + RH98 + Elevation+ Precipitation |
| Simpson | FHD + RDVI + RH98 + Slope |
| Richness | Elevation + FHD + Precipitation + RDVI |

The order ranking represents the importance and ranking of the variables whose impact is shown in Figure 9.

The variable selection method revealed that among the 15 candidate variables, two of them were selected for all models. FHD was the GEDI variable selected for all diversity models, while RDVI was the spectral index from Sentinel-2 also selected in all models. Additionally, the variable selection method for the Shannon and Simpson models revealed that GEDI metrics were more important in relation to other data sources. For these models, FHD and RH98 were selected to estimate species diversity in the Cerrado. This shows that for the Simpson and Shannon models, GEDI variables represented 50% and 40% of the total selected variables, respectively. The auxiliary variables, Elevation (SRTM) and Precipitation (ERA-5), also proved to be important for diversity models, as they were selected in two of the three models developed (i.e: Shannon and Richness).

In general, topographic variables also proved to be relevant for variable selection. All models included at least one topographic variable. Terrain elevation was selected for the Shannon and Richness models, while terrain slope proved to be important for the Simpson model. For the five Sentinel-2 candidate spectral index variables, only RDVI was selected for the species diversity estimation in Cerrado. RDVI was selected for all models. The OOB error values decreased continuously as the variable number increased from 0 to 4 for Simpson and Richness models and increased from 0 to 5 for Shannon’s model (Figure 9).

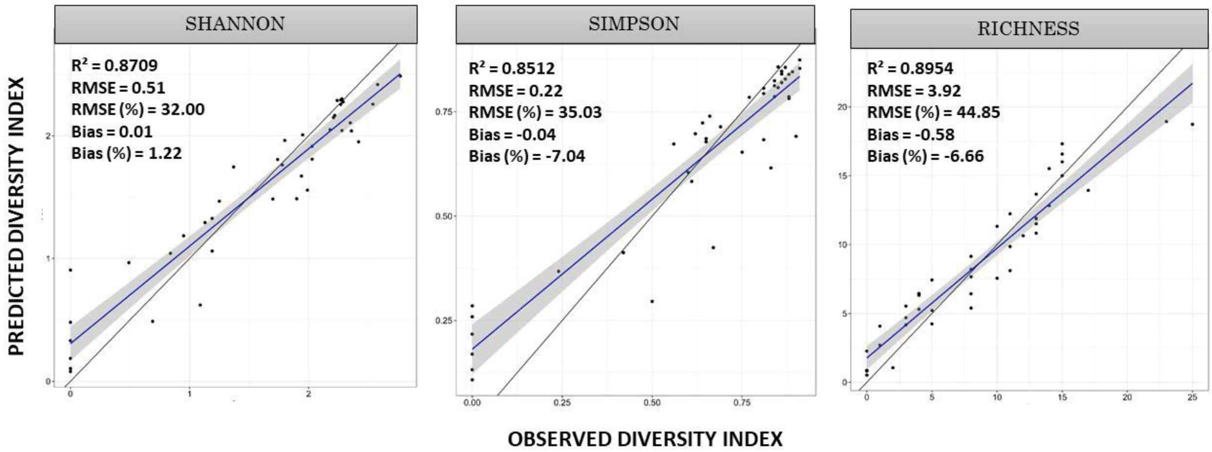
FIGURE 9 - VARIABLE SELECTION USING RANDOM FORESTS (VSURF) GIVING THE NUMBER OF VARIABLES THAT MEET THE REQUIREMENTS FOR PREDICTION.




## 6.3 PREDICTIVE MODELS FOR SPECIES DIVERSITY INDICES

Overall, all models performed well during training with $R^2 > 0.858$, RMSE <45%, Bias < -7.94 (FIGURE 10). Among the three models, Richness showed the best model fit with an $R^2$ of 0.89, Shannon had slightly lower performance with an $R^2$ of 0.87, while Simpson had the lowest performance with an $R^2$ of 0.85. The Shannon and Richness indices were more accurately estimated with models, producing $R^2$ values of 0.52 and 0.56, RMSE of 36% and 54% in validation, respectively (TABLE 8). On the other hand, the model that estimates the Simpson index showed low performance during validation ($R^2 = 0.24$), despite the RMSE % of this model in validation being the lowest (35.29%) among the others.

FIGURE 10 - TRAINING RESULTS FOR ESTIMATING THE DIVERSITY INDICES (SHANNON, SIMPSON AND RICHNESS) USING RANDOM FOREST, GEDI WAVEFORM METRICS AND PASSIVE OPTICAL IMAGING AS PREDICTORS. $R^2$ = COEFFICIENT OF DETERMINATION; RMSE = ROOT MEAN SQUARE ERROR; AND BIAS.




The models tended to overestimate the values for diversity indices in Cerrado. The species richness model had an RMSE of 5.03, while Shannon and Simpson had RMSEs of 0.63 and 0.24, respectively for validation. However, all models had low bias values. Shannon and Simpson both had a bias value of 0.07, while Richness had a value of -0.56.

TABLE 8 - CROSS-VALIDATION PERFORMANCE ASSESSMENT IN 500 ITERATIONS OF MODELS USED TO ESTIMATE THE DIVERSITY INDEX (SHANNON, SIMPSON AND RICHNESS) USING RANDOM FOREST, GEDI WAVEFORM METRICS AND CONVENTIONAL PASSIVE OPTICAL IMAGING AS PREDICTORS. $R^2$ = COEFFICIENT OF DETERMINATION; RMSE = ROOT MEAN SQUARE ERROR; BIAS, AND T.

| Diversity Index | $R^2$ | RMSE | RMSE% | Bias | Bias % | t |
|---|---|---|---|---|---|---|
| Shannon | 0.52 | 0.63 | 37.47% | 0.07 | 4.44 | 0.76 |
| Simpson | 0.24 | 0.24 | 35.29% | 0.07 | 10.2 | 0.31 |
| Richness | 0.56 | 5.03 | 53.89% | -0.56 | -6.02 | 0.71 |

## 6.4 DIVERSITY INDEX CHARACTERIZATION ACROSS THE CERRADO BIOME

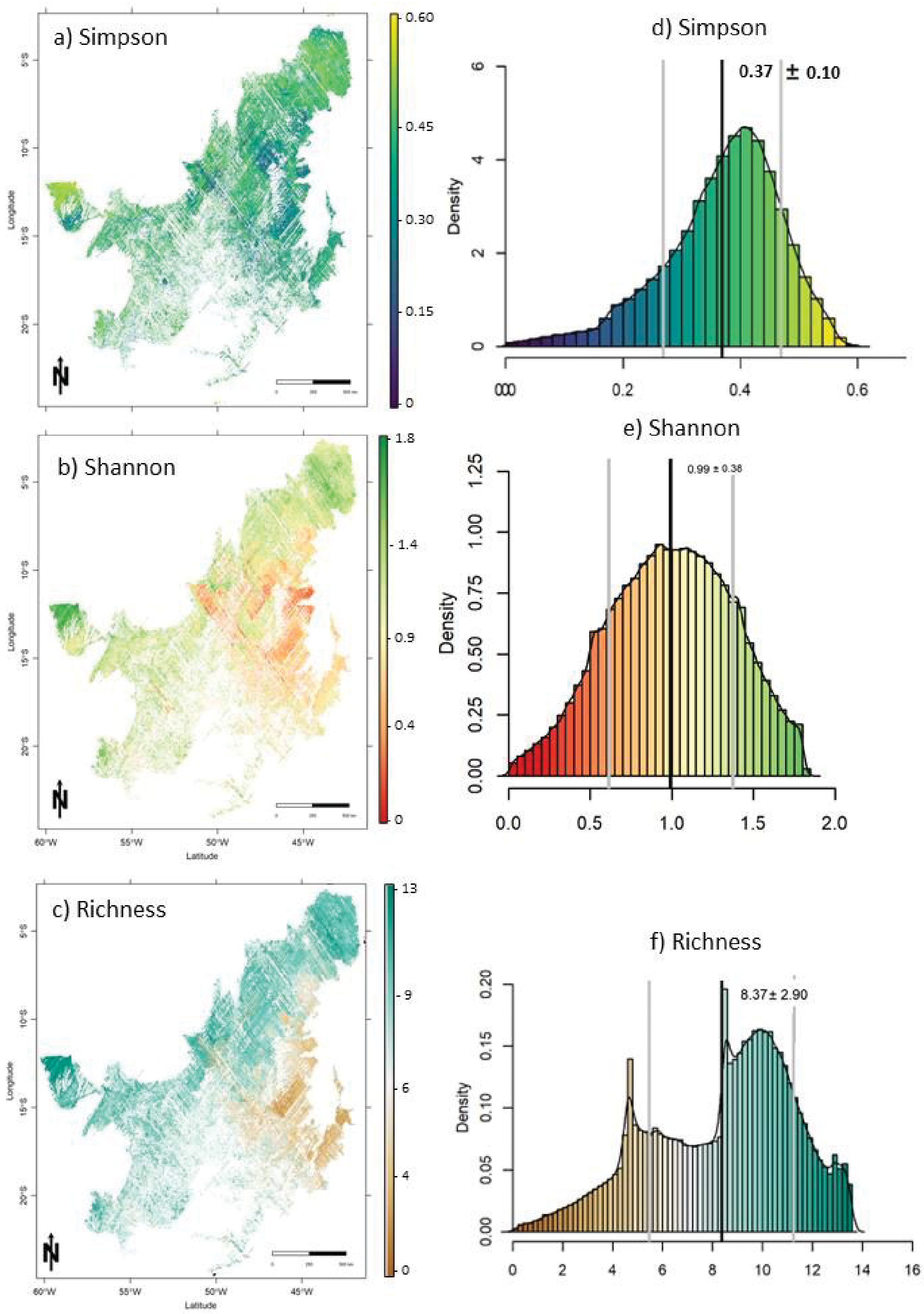
Species diversity indices estimates were obtained by applying the models to the in-orbit GEDI data stacked with optical data (Sentinel-2, ERA-5, SRTM).

Estimates were obtained for the entire Cerrado Biome using GEDI footprints with a radius of 25 m. The spatial variation of estimates of species diversity indices in the Cerrado is shown in Figure 11. These maps allowed us to identify regions of the Cerrado with greater species diversity (e.g., 11°S ~ 58°W Fig. 11i) and locations with lower diversity (e.g., 11°S ~ 44°W Fig. 11i). The distribution of estimates was mostly evenly distributed for all indices (FIGURE 11 d-f). The estimated mean values of Shannon, Simpson, and Richness were 0.99 ± 0.38, 0.37 ± 0.10, and 08.37 ± 2.90, respectively.

Based on the diversity indices estimates for the entire biome, we propose a classification to indicate areas of low, medium and high diversity. In this sense, for the Simpson index as shown in Figure 11 d, the distribution obtained indicates that values obtained up to 0.27 can be considered as low species diversity. Values between 0.27 and 0.47 indicate medium species diversity, and values greater than 0.47 indicate high species diversity values in the Cerrado using the Simpson index.

For the Shannon index as shown in Figure 11 e, the distribution obtained indicates that values obtained up to 0.61 can be considered as low species diversity. Values between 0.61 and 1.37 indicate an average species diversity, and values greater than 1.37 indicate high species diversity values in the Cerrado using the Shannon index. Due to the bimodal distribution obtained for Richness, we will not consider the classification for this index.

FIGURE 11 - LARGE SCALE DIVERSITY INDICES PREDICTION MAPS (A-C) AND DISTRIBUTION (D-F) AT THE 1 KM OF SPATIAL RESOLUTION FOR THE ENTIRE CERRADO BIOME.




Diversity index maps show regions with high, medium and low species diversity in the Brazilian Cerrado. It is possible to observe that all maps show the same distribution pattern. That is, in the north and east of the Cerrado it is possible to observe values that represent high diversity according to the model estimates. Further to the south, center and west of the biome, there is a greater arrangement of values that represent a low species diversity. The uncertainty of the predictions was similarly distributed across Cerrado (FIGURE 12), with a pattern of lower uncertainty in regions with more GEDI footprints (FIGURE 12 d-f).

Maps of species diversity index estimates were also crossed with information from highways and urbanized areas. With this information, we realize that there is a greater predisposition for areas with higher values of species diversity, in places where there is a greater distance from large centers and major intersections between highways (FIGURE 13 a-c). For example, towards the center of the biome we have large cities such as Belo Horizonte, Brasília and Goiás, which represent the largest cities contained within the limits of the biome. The areas with higher values of species diversity can be observed in the left region from east to north. In these regions it is possible to see few urbanized areas and fewer highway crossings.

Another analysis that we were able to observe is the distribution of Conservation Units (CU) contained in the Cerrado biome (FIGURE 14 a-c). A first point is the low number of Conservation Units in the biome, in total there are 51, divided into 29 for sustainable use and 22 for integral protection. Another important point is the uniform distribution and the absence of units in certain locations. For example, to the east of the biome on the border between Rondônia and Mato Grosso, we have one of the places with the highest values of species diversity, however, there is no CU there or nearby.

FIGURE 12 - LARGE SCALE DIVERSITY INDICES UNCERTAINTY PREDICTION MAPS (A1-C1) AND DISTRIBUTION (A2-C2) AT THE 1 KM OF SPATIAL RESOLUTION FOR THE ENTIRE CERRADO BIOME.

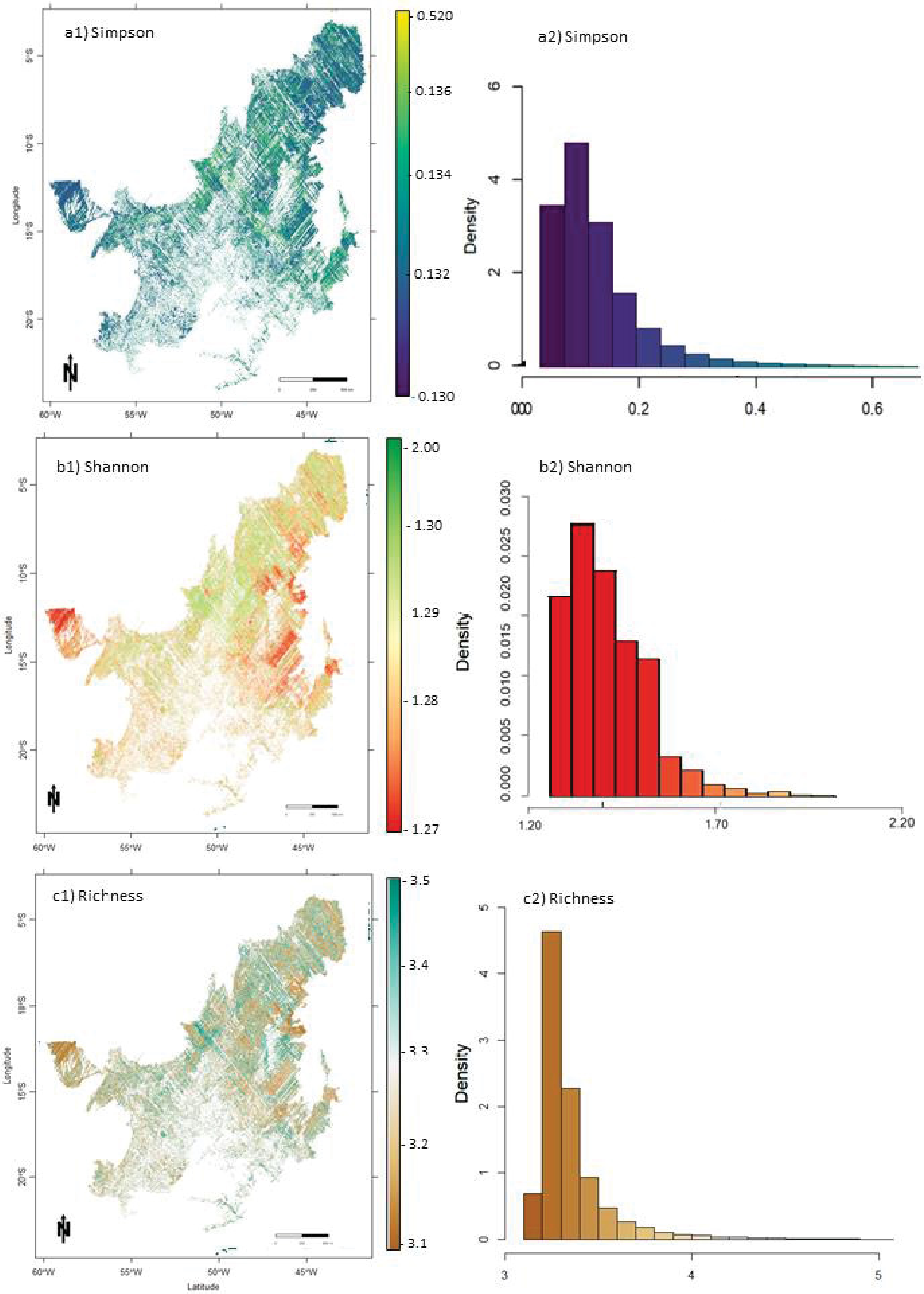


FIGURE 13 - LARGE SCALE DIVERSITY INDICES UNCERTAINTY PREDICTION MAPS (A1-C1) AND DISTRIBUTION (A2-C2) AT THE 1 KM OF SPATIAL RESOLUTION FOR THE ENTIRE CERRADO BIOME.

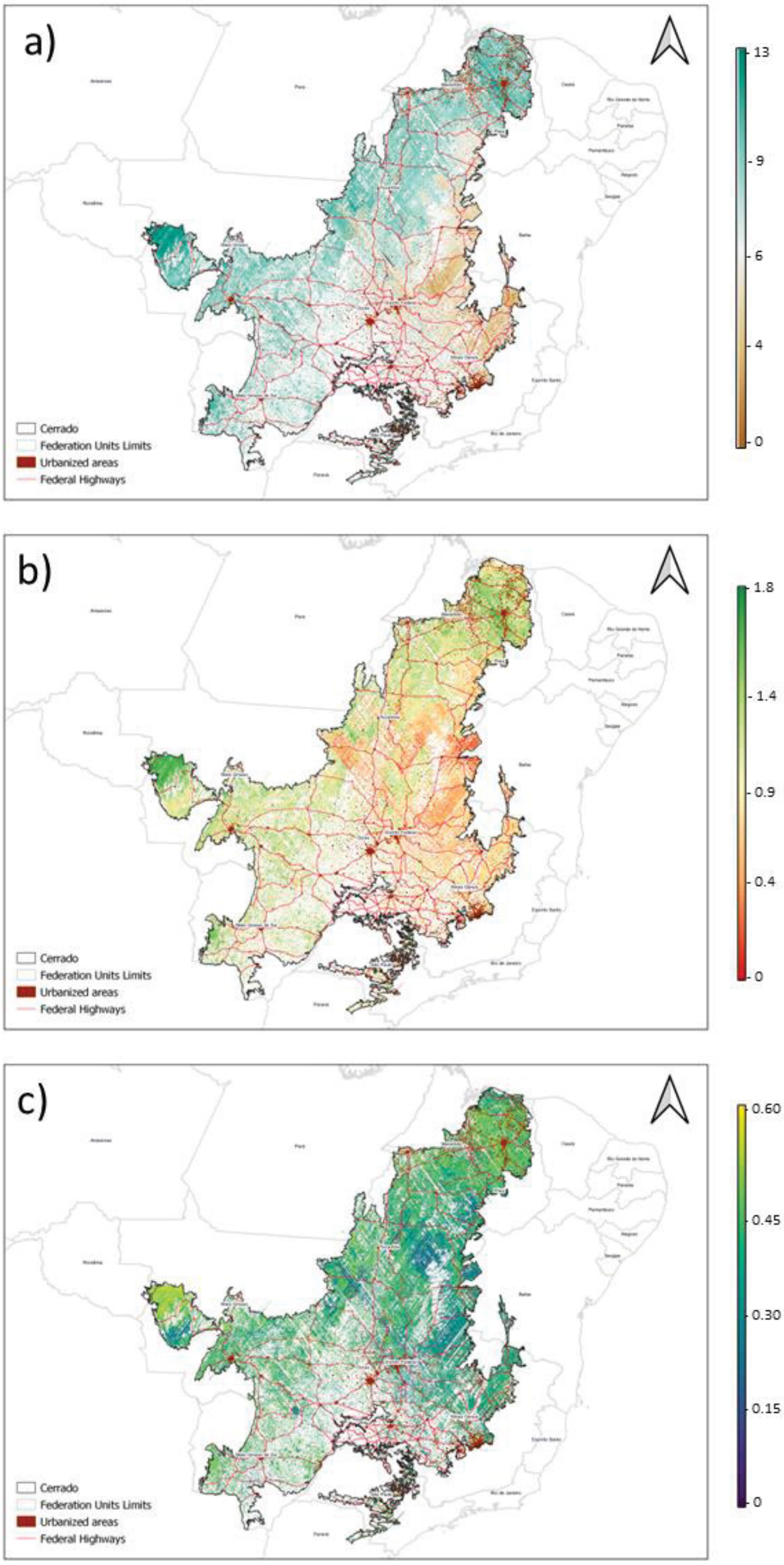
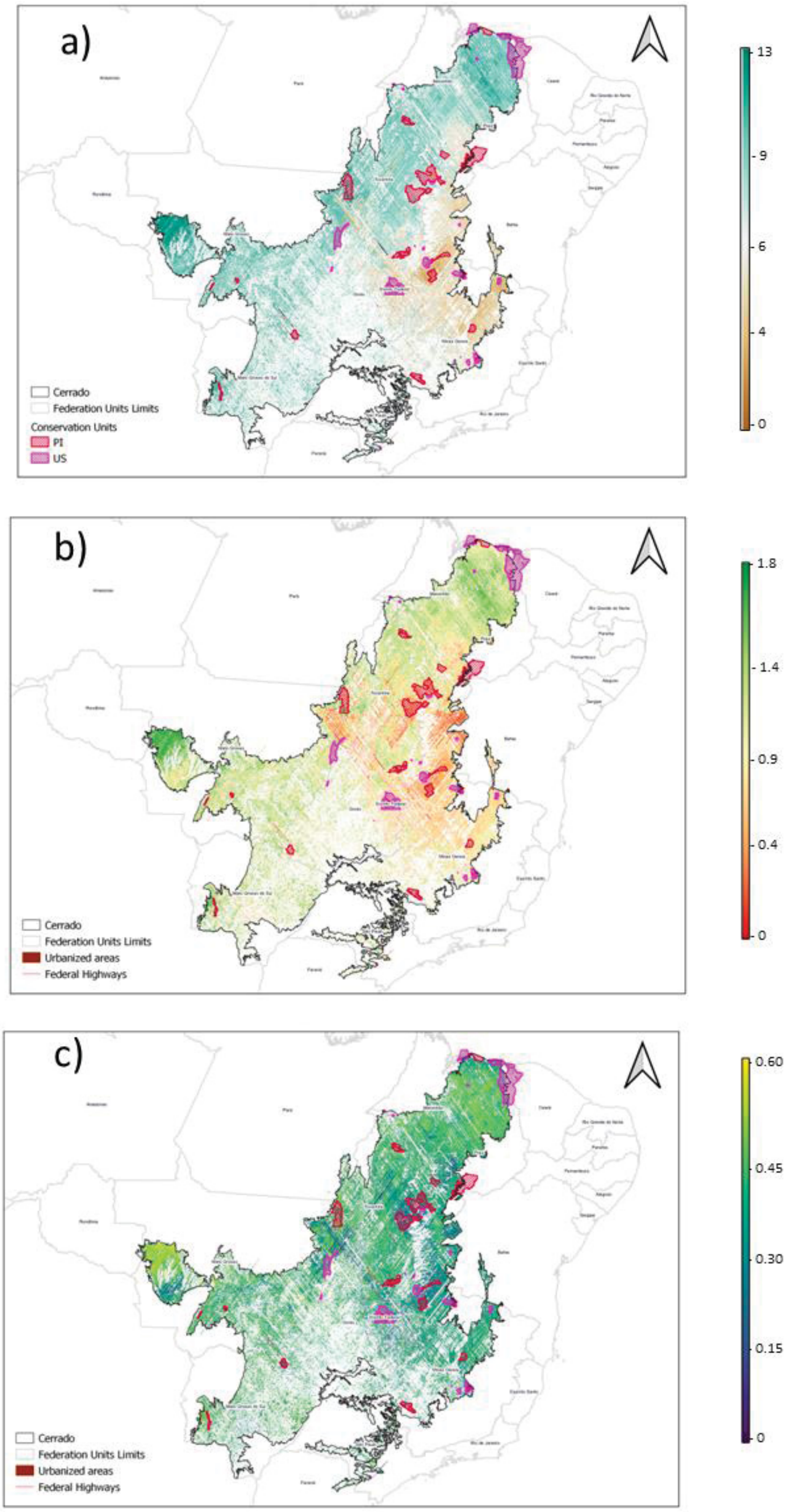


FIGURE 14 - LARGE SCALE DIVERSITY INDICES UNCERTAINTY PREDICTION MAPS (A1-C1) AND DISTRIBUTION (A2-C2) AT THE 1 KM OF SPATIAL RESOLUTION FOR THE ENTIRE CERRADO BIOME.



## 7 DISCUSSION

In this study, we were able to estimate large-scale forest diversity for the Brazilian tropical savanna (Cerrado), using GEDI data combined with conventional passive optical imagery from space. This study represents a first step to understand the relationship between tree species diversity and the variability of multisource remote sensing data in the Cerrado. To date, no similar studies have been carried out in this ecosystem, and this suggests promising potential in using free satellite-derived structural and optical data in combination with machine learning to understand tree species diversity of the Cerrado.

The spatial predictions of tree species diversity are scarce for savanna ecosystems. Our results demonstrated that the combined use of GEDI and conventional passive optical imagery data can improve large-scale species diversity indices estimates. According to Oldeland et al. (2010) descriptors of alpha diversity, such as Shannon, which is less affected by rare species than Richness, would be more readily predicted by remote sensing data. This aligns with our findings, however, Shannon and Richness were more accurate than Simpson. The Simpson index gives more weight to species with higher proportions while Richness is only based on species presence/absence. In our models' case, adding GEDI variables may have contributed to species richness estimates, since the presence of rare species in the understory influences species richness but would be difficult to detect by passive sensors. In contrast, dominant species are more detectable upon remote sensing perspective. These findings suggest that multi-source models combining structural and optical data can be used to map large-scale tree species diversity in Cerrado.

A knowledge of the variables that contributed most to model accuracies is important in modeling. It helps to select robust key variables, reduces redundancy and noise in the prediction and characterization of vegetation attributes (Millard and Richardson, 2015). Regarding the variables selected by the VSURF algorithm, the Foliage Height Diversity (FHD) proved to be considerably important for all diversity models. FHD was selected as the most important variable for Simpson, and the second most important variable for Shannon and Richness. The critical role of FHD normally indicates a more complex forest structure, which is interesting for ecosystems such as savannas that have complex stem and canopy structures. Structural differences between tree species provide a different directional gap probability, which underlies LiDAR-based forest diversity estimates. Direct correlations between tree species diversity by indices and FHD derived from GEDI have been confirmed. RH98 also proved to be an important variable for estimating large-scale diversity indexes in Cerrado. Canopy height information has been evaluated as an important variable for modeling forest parameters, such as Biomass (Duncanson et al., 2020; Dorado-Roda et al., 2021; Crockett et al., 2023); Fuel load (Leite et al., 2022); Forest Volume (Chen et al., 2021) and even in studies that estimated species diversity (Simonson et al., 2012, Hakkenberg et al., 2023). In this way, GEDI metrics will become one of the most important parameters in estimating tree species diversity at large scale in Cerrado. Additionally, the vertical structure diversity captured by GEDI may also contribute to fauna diversity, providing greater diversity of habitats (open fields, tree canopies, under canopies).

Our findings also showed the importance of the spectral information revealed here by RDVI from the Sentinel-2. The RDVI was strongly associated with all tree species diversity indices derived from field data. This index combines the benefits of NDVI and Difference Vegetation Index (DVI), performing well in both dense canopy and sparse vegetation (i.e., canopy gap) conditions (Roujean and Breon, 1995). All vegetation indices tested in this study cover the NIR and red electromagnetic spectrum region, which are wavelength regions often defined as the most relevant for studying differences in vegetation structures (Chuvieco, 2002; Vaiphasa et al., 2007). Previous studies have found that the use of VIs is more correlated with abundance indices (Shannon and Simpson) (Meng et al., 2016; Madonsela et al., 2017), but they can also partially explain Richness (Madonsela et al., 2017).

Furthermore, Castillo-Santiago et al. (2010) reported that spectral information is better used to explain the variation in forest structure at lower biomass levels. This finding may have contributed to our results, while da Costa et al. (2021) for the same study area (Cerrado) found mean aboveground biomass estimates ranging from 21.28 to 99.35 $Mg.ha^{-1}$. The RDVI has a measurement scale that ranges from 0 to well beyond 1, and such an open scale facilitated its ability to explain tree species diversity in our study. Most studies report NDVI as a potential variable for forest species diversity. However, the scale problem of the NDVI has long been recognized as limiting its ability to detect forest canopy variation (Huete et al., 2002). Therefore, it is not surprising that the RDVI has greater explanatory power than the other indices.

Interestingly, topographic data sources were also important for the generated models. Elevation was important for the Shannon and Richness models, while Slope was necessary for the Simpson model. The importance of elevation and slope has also been highlighted in studies relating to the tree species classification (Liu et al., 2023) as they can reflect elevation zones and differences in exposure to solar radiation (Ren et al., 2009). Previous studies claim elevation was one of the most important metrics for estimating tree (Robinson et al., 2018; Read et al., 2020; Grabska et al., 2020). This statement aligns with the result of the Richness model, since elevation was the most important variable for this model. These studies explored the relationship between vegetation and topographic factors and found that, with an increase in elevation and aspect, vegetation growth rates also increased. In addition, as Morris et al. (2014) reported, not all diversity indices are equally correlated with landscape parameters. For example, species richness correlates better with landscape parameters than Shannon's index. This is partially in line with our results, in which elevation was important not only for the Richness model, but also for the Shannon model. These results reinforce those from other studies where remotely derived abiotic factors related to topographic and edaphic properties were found to be significant predictors of tree species richness (Fricker et al., 2015; Zellweger et al., 2016; Hakkenberg et al., 2018). Finally, previous research has found that precipitation also has a strong effect on woody species richness with strongly corroborates with our results (Richerson et al., 1980, Li et al., 2019. Once again, the precipitation variable was highlighted for Shannon and Richness indices to compose the models. So, this suggests that remote sensing explains the variation in species diversity better when integrated with environmental variables, since they are also known to influence spatial patterns of natural resources (Silva et al., 2017; Shoko et al., 2019).

Although our models have proven to be consistent in estimating and mapping large-scale species diversity indices in Cerrado, this study has some limitations, and the methodology can be improved in the future. First, we used a limited number of plots (50). Our results could be significantly better with a more robust database for training and validating our models, although we recognize that collecting such data is expensive and time-consuming. Additionally, our efforts to cover as many vegetation types as possible in Cerrado were also limited due to the size and logistics of access to the field in some regions. As discussed earlier, factors such as terrain topography and climatic conditions directly influence species diversity, while work has shown that structural profiles generated from GEDI data are capable of capturing patterns of biodiversity in forest ecosystems without ancillary data on topography or climate (Hakkenberg et al., 2023), it is uncertain if this approach holds for savanna systems like the Cerrado, thus more data are needed to constrain the uncertainty in modeling these relationships. Second, the relationship between the spectral reflectance of the tree species diversity could vary among different seasons; Torresani et al. (2019) highlighted the importance of a time-series approach to estimate tree species diversity, as their results showed that the relation between field data and spectral heterogeneity indices varies strongly within a year. Mandosela et al. (2017), using two Worldview-2 images captured during peak productivity and the senescence season in a South African Savanna, identified the latter as the most appropriate for tree species diversity modeling. Arekhi et al. (2017), examined the relationship between tree species diversity (Shannon index) and several Landsat-8 TM spectral bands in a temperate forest in Turkey, and identified that the image acquired in late June facilitated the development of the model more accurately. Herein, images from May and August were used to obtain median composites and thus obtain the VIs. This procedure was carried out to ensure that our data was free of cloud and cloud shadows, so we expanded it a month before and a month after the field campaign, referring to a transition between autumn-spring but still within the dry season. In addition, spectral resolution plays an important role in understanding landscape diversity, resulting in an increase in accuracy of diversity estimation by adding spectral wavelengths (Rocchini, 2007). Khare et al., (2019) with RapidEye (including the red-edge band) found that overall, Rao's Q mean values for all sample plots were higher. Third, the plot design could be improved in the future. Hernández-Stefanoni et al. (2014) proved that in the case of species richness, that it is much more efficient to capture local landscape variability using a cluster of plots.

Habitat heterogeneity has often been associated with species richness. Several authors have reported strong associations between spectral heterogeneity (as a proxy for habitat heterogeneity) and species richness (Hernández-Stefanoni et al., 2012; Levin et al., 2007). Future studies could be concerned with forming a more robust database by increasing the number of plots in each vegetation type and considering different regions. Furthermore, some research may address whether there are significant differences in model predictions caused by seasonality, since the results of research carried out in forest formations belonging to distant biomes and subject to different climate regimes can be quite contrasting (Chrysafis et al., 2020). It would also be interesting to use new approaches to increase the quantity and coverage of canopy structure measurements by GEDI, since the mission was programmed to collect traces. Future studies should also focus on other multi-spectral vegetation diversity indices, including those extracted from red-edge bands, since they enable recording even subtle differences in leaf structure and chlorophyll content (Delegido et al., 2011, Schuster et al., 2012).

Statistically rigorous methods for communicating results quickly to the scientific community, government, industrial stakeholders and the public are a critical element of successful biodiversity integrity monitoring programs (Noss, 1990; Debinski and Humphrey, 1997; Yoccoz et al., 2001; Nichols and Williams, 2006; Lovett et al., 2007). Understanding the tree species diversity status in biodiversity conservation is crucial as it provides management with the necessary baseline information about tree species distribution in the ecosystem, which is essential in planning and management. This information is of great importance for the Cerrado, which has an intense predatory exploitation: countless animals and plants are at risk of extinction and only 8.21% of the territory's total area is legally protected by conservation units. As suggested by Marselis et al. (2019), we also believe that future research should focus on the transferability of the structure-diversity models to other regions and continents to establish the potential of using this method in various biomes. Additionally, the information derived from our approach may open new opportunities for future studies of global importance. Topics such as the assessment of biodiversity loss in ecosystems and pandemic effects (e.g., COVID-19) are extremely important and could be explored better.

## 8 CONCLUSION

We developed a new workflow for large-scale tree species diversity mapping for the Brazilian tropical savanna (Cerrado) combining GEDI, optical and environmental data. Three multi-source models were developed to estimate large-scale tree species diversity indices in the Cerrado. The Shannon and Richness indices were more consistent than the Simpson Index for evaluating tree species diversity in the Cerrado. GEDI metrics can capture information related to vegetation structures and significantly improve the accuracy of tree species diversity estimates and mapping. The RDVI vegetation index was selected for all models, confirming our hypothesis that spectral information contributes to the description of species diversity. The Elevation and Slope variables were also necessary for composing the models, while the Precipitation was more favorable for predicting Shannon and Richness indices. Remote sensing provides systematic spatial and temporal data on vegetation attributes, which can be assessed on a large scale. Our findings will contribute to identifying locations with high and low species richness, and abundance of shrubs and trees, through remote sensing data. The maps provided in this study will be valuable for the assessment and management of tree species diversity in the Cerrado biome.

## 9 RECOMMENDATIONS

It is recommended to expand field collection to obtain a more robust database with greater variability of regions of the Cerrado biome, and not just of different phytophysiognomies. Furthermore, new research may address whether there are significant differences in model predictions caused by seasonality. Another perspective would also be to evaluate the inclusion of other variables, mainly GEDI. In this study, we pre-select candidate variables before any variable selection method. In this sense, other variables can contribute to improving estimates of species diversity indices in the Brazilian Cerrado. The same is valid for variables from the Sentinel-2 satellite's MSI sensor. In this study, we only focus on the spectral information of RED and NIR wavelengths. Finally, multi-temporal research would be interesting to assess whether there are areas in the process of losing diversity and whether there are areas in the process of vegetation restoration and consequently gains in local/regional diversity.

## ACKNOWLEDGMENTS

This project was supported by the Brazilian National Council for Scientific and Technological Development (CNPq, grant 442640/2018-8, CNPq/Prevfogo-Ibama N° 33/2018). D. R.A.A. was supported by the São Paulo Research Foundation (#2018/21338-3). A.C. was financed in part by MCTIC/CNPq N° 28/2018

(#408785/2018-7; #438875/2018-4), CNPq N° 09/2018 (#302891/2018-8). A.P.D.C. was supported by CNPq #88887.373249/2019-00, MCTIC/CNPq Nº 28/2018 (#408785/2018-7; #438875/2018-4) and CNPq Nº 09/2018 (#302891/2018-8). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. M.E.F. is a CNPq Research Fellow (grant #315699/2020-5). We gratefully acknowledge the following undergraduate and graduate students involved in the field work campaign: Alberto A. Gontijo e Silva, Alexandre S. C. Filho, André Felipe C. Lima, Bernardo dos S. de Almeida, Carlos Magno M. de Oliveira, Gilberto do A. Pacheco; Gustavo R. Lattanzi, Iago Henrique F. da Silva, Irene M. Barbosa, Ivo S. Moreira, Jacson A. A. Machado, Jean Victor N. Paiva, Junia S. M. Macedo, Leandra Dietrich, Lídia A. de Aguiar, Matheus Gunther M. Soares, Nelson Amaral, Nivaldo R. J. Junior, Reginaldo Arthur G. Marcelino, Thiago Trajano. Moreover, we thank the park managers: Edward Elias Junior, Celso Lago-Paiva and Leandro Chagas from the Serra do Cipó National Park, Maria Carolina A. Camargos from the Chapada dos Veadeiros National Park and Renato Diniz Dumont from the Paraopeba National Forest. GatorEye data collection and processing by AMAZ and ENB was supported in part by the McIntire-Stennis program of the USDA, and the School of Forest Resources and Conservation and the Center for Latin America Studies at UFL. We thank Trina Merrick for assisting with the GatorEye UAV-lidar system repair. CA was supported by The University of Colorado Boulder Grand Challenge, CIRES, Earth Lab. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA of U.S. Government determination or policy.

## REFERENCES

ADÁMOLI, J; MACEDO, J; AZEVEDO, L.G; NETTO, J.M. Caracterização da região dos cerrados. *In:* GOEDERT, W.J; (ed). **Solos dos cerrados: tecnologias e estratégias de manejo**. Planaltina: - Embrapa-CPAC, São Paulo: Nobel, 1987. p. 33-98.

ALMEIDA, D. R; STARK, S. C; VALBUENA, R; BROADBENT, E. N; SILVA, T. S; DE RESENDE, A. F; CARDIL, A; SILVA, C. A; AMAZONAS, N; ZAMBRANO, A. M. A; BRANCALION, P H. S. A new era in forest restoration monitoring. **Restoration Ecology**, v. 28 (1), p. 8-11, 2020.

AREKHI, M; YILMAZ, O.Y; YILMAZ, H; AKYÜZ, Y.F. Can tree species diversity be assessed with Landsat data in a temperate forest? **Environmental Monitoring and Assessment**, v. 189, p. 1-14, 2017.

BALTSAVIAS, E. P. Airborne laser scanning: basic relations and formulas. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 54, n. 2-3, p. 199-214, 1999.

BATISTA, A. P. B; MELLO, J. M. D; RAIMUNDO, M. R; SCOLFORO, H. F; REIS, A. A. D; SCOLFORO, J. R. S. Species richness and diversity in shrub savanna using ordinary kriging. **Pesquisa Agropecuária Brasileira**, v. 51, p. 958-966, 2016.

BECK, J; ARMSTON, J; HOFTON, M; LUTHCKE, S. **Global Ecosystem dynamics investigation (GEDI) level 1B user guide**. Document version 2.0. USGS Earth Resources Observation and Science (EROS) Center: NASA's Land Processes Distributed Active Archive Center (LP DAAC), https://lpdaac. usgs. gov/documents/590/GEDIL01_User_Guide_V1. pdf, 2020. Accessed: 13 March 2021.

BELGIU, M. AND DRĂGUŢ, L. Random Forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24-31, 2016.

BENDINI, H. D. N; FONSECA, L. M. G; SCHWIEDER, M; RUFIN, P; KORTING, T. S; KOUMROUYAN, A; HOSTERT, P. Combining environmental and Landsat analysis ready data for vegetation mapping: a case study in the Brazilian savanna biome. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 43, p. 953-960, 2020.

BENITES, V.M; CAIAFA A.N; MENDONÇA, E.S; SCHAEFER C.E.G.R; KER, J.C. (2003). Solos e vegetação nos complexos rupestres de altitude da Mantiqueira e do Espinhaço. **Floresta e Ambiente**, v. 10, p. 76-85, 2023

BLAIR, J. B; AND HOFTON, M. A. Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. **Geophysical Research Letters**, v. 26, n. 16, p. 2509-2512, 1999.

BOLAND, J; T. AGER, E. EDWARDS, E. FREY, P. JONES, R.K. JUNGQUIET, A.G. LAREAU, J. LEBARRON, C.S. KING, K. KOMAZAKI, C. TOTH, S. WALKER, E. WHITTAKER, P. ZAVATTERO, H. ZUEGGE, Cameras and sensing systems. **Manual of Photogrammetry**, v. 8, p. 581-676. 2004.

BORLAUG, N. E. Feeding a world of 10 million people: the miracle ahead. **Plant Tissue Culture and Biotechnology**, v. 3, p. 119-127, 1997.

BOUCHER, P.B; HANCOCK, S; ORWIG, D.A; DUNCANSON, L; ARMSTON, J; TANG, H; SCHAAF, C; Detecting change in forest structure with simulated GEDI lidar waveforms: a case study of the hemlock woolly adelgid (HWA; *Adelges tsugae*) infestation. **Remote Sensing**, v. 12, n. 8, p. 1304, 2020.

BRAZIL. **MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION**. Third national communication of Brazil to the United Nations framework convention on climate change. p. 333 2016.

BRASIL, Ministério do Meio Ambiente. Mapeamento do uso e cobertura do Cerrado: Projeto TerraClass Cerrado.2015. Disponível em: http://www.dpi.inpe.br/tccerrado/download.php. Acesso em: 01 abr. 2023.

BREIMAN, L. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.

BROADBENT, E.N; ALMEYDA ZAMBRANO, A.M; OMANS, G; ADLER, B; ALONSO, P; NAYLOR, D; CHENEVERT, G; MURTHA, T; PRATA, G; DE ALMEIDA, D.R.A; ET AL; **The GatorEye Uninhabited Flying Laboratory: Sensor Fusion for 4D Ecological Analysis through Custom Hardware and Algorithm Integration**. Available online: www.gatoreye.org. Accessed: 5 May 2021.

BROWER, J. E; ZAR, J. H. **Field and laboratory methods for general ecology**. 2ª ed. Dubuque: Wm. C. Publishers, 1984.

CABACINHA, C. D; DE CASTRO, S. S. Relationships between floristic diversity and vegetation indices, forest structure and landscape metrics of fragments in Brazilian Cerrado. **Forest Ecology and Management,** v. 257, n. 10, p. 2157-2165, 2009.

CALIXTO JÚNIOR, J. T. **Análise estrutural de duas fitofisionomias de caatinga em diferentes em diferentes estados de conservação no semiárido pernambucano**. 2009. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Campina Grande, Patos - PB. 2009.

CARNEIRO, L. S. **Uso de geotecnologias para mapeamento da peroba-rosa (Aspidosperma polyneuron Müll. Arg)**. 26f. Trabalho de conclusão de curso - Universidade Federal Rural Do Rio De Janeiro. Rio de Janeiro. 2023.

CASTILLO-SANTIAGO, M.A; RICKER, M; DE JONG, B.H. Estimation of tropical forest structure from SPOT-5 satellite images. **International Journal of Remote Sensing**, v. 31, n. 10, p. 2767-2782, 2010.

CENTENO, J. A. S. **Sensoriamento remoto e processamento de imagens digitais**. Curitiba: UFPR**,** v. 219, 2003.

CERQUEIRA, E. B. **O Cerrado e suas comunidades tradicionais: Impactos e a percepção estudantil relacionada a esses aspectos**. 122 p. Dissertação de Mestrado – Campus Urutaí, Instituto Federal Goiano, Urutaí, 2023.

CHEN, L; REN, C; ZHANG, B; WANG, Z; LIU, M; MAN, W; AND LIU, J. Improved estimation of forest stand volume by the integration of GEDI LiDAR data and multi-sensor imagery in the Changbai Mountains Mixed forests Ecoregion (CMMFE), northeast China. **International Journal of Applied Earth Observation and Geoinformation**, v. 100, p. 102326, 2021.

CHRYSAFIS, I; KORAKIS, G; KYRIAZOPOULOS, A. P; AND MALLINIS, G. Predicting tree species diversity using geodiversity and Sentinel-2 multi-seasonal spectral information. **Sustainability**, v. 12, n. 21, p. 9250, 2020.

COLWELL, R. K. Biodiversity: concepts, patterns, and measurement. **The Princeton Guide to Ecology**, v. 663, p. 257-263, 2009.

COSTANZA, R; DE GROOT, R; SUTTON, P; VAN DER PLOEG, S; ANDERSON, S.J; KUBISZEWSKI, I; FARBER, S. AND TURNER, R.K; Changes in the global value of ecosystem services. **Global Environmental Change**, v. 26, p. 152-158, 2014.

CROCKETT, E.T.H; ATKINS, J.W; GUO, Q; SUN, G; POTTER, K.M; OLLINGER, S; SILVA, C.A; TANG, H; WOODALL, C.W; HOLGERSON, J; XIAO, J. Structural and species diversity explain aboveground carbon storage in forests across the United States: Evidence from GEDI and forest inventory data. **Remote Sensing of Environment**, v. 295, p. 113703, 2023.

DA COSTA, M. B. T; SILVA, C. A; BROADBENT, E. N; LEITE, R. V; MOHAN, M; LIESENBERG, V; STODDART, J; DO AMARAL, C. H; DE ALMEIDAI, D R A; DA SILVA, A L; GOYA, L. R. R. Y; CORDEIRO, V. A; REX, F.E,, HIRSCH, A; MARCATTI, G. E; CARDIL,A; DE MENDONÇA, B. A. F; HAMAMURA, C; CORTE, A.P; MATRICARDI, E. A. T; KLAUBERG, C. Mapping total aboveground biomass density in the Brazilian savanna using high-density UAV-lidar data**. Forest Ecology and Management**, v. 491, p. 119155, 2021.

DAMASCO G; FONTES C; FRANÇOSO R; HAIDAR R. The Cerrado biome: a forgotten biodiversity hotspot. **Frontiers for Young Minds**, v. 6, 2018.

DE SOUZA MENDES, F; BARON, D; GEROLD, G; LIESENBERG, V; ERASMI, S. Optical and SAR remote sensing synergism for mapping vegetation types in the endangered Cerrado/Amazon ecotone of Nova Mutum—Mato Grosso. **Remote Sensing,** v. 11, n. 10, p. 1161, 2019.

DEBINSKI, D. M. An integrated approach to biological diversity assessment. **Natural Areas Journal**, v.17, p. 355-363. 1997.

DELEGIDO, J; VERRELST, J; ALONSO, L; AND MORENO, J. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. **Sensors**, v. 11, n. 7, p. 7063-7081, 2011.

DONG, P; CHEN, Q. **LiDAR remote sensing and applications**. Boca Raton, FL: Taylor Francis, 210p. 2018.

DORADO-RODA, I; PASCUAL, A; GODINHO, S; SILVA, C. A; BOTEQUIM, B; RODRÍGUEZ-GONZÁLVEZ, P; GONZALEZ FERREIRO, E. M; GUERRA-HERNÁNDEZ, J. Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in Mediterranean forests. **Remote Sensing**, v. 13, n. 12, p. 2279, 2021.

DRUSCH, M; DEL BELLO, U; CARLIER, S; COLIN, O; FERNANDEZ, V; GASCON, F; HOERSCH, B; ISOLA, C; LABERINTI, P; MARTIMORT, P; MEYGRET, A; SPOTO, F; SY, O; MARCHESE, F; BARGELLINI, P. Sentinel2: ESA's optical high-resolution mission for GMES operational services. **Remote Sensing of Environment**, v. 120, p. 25–36, 2012.

DUBAYAH, R. O; DRAKE, J. B. Lidar remote sensing for forestry. **Journal of Forestry**, v. 98, n.6, p. 44-46. 2000.

DUBAYAH, R; BLAIR, J. B; GOETZ, S; FATOYINBO, L; HANSEN, M. HEALEY, S; HOFTON, M; HURTT, G; KELLNER, J; LUTHCKE, S; ARMSTON, J; TANG, H; DUNCANSON, L; HANCOCKM, S; JANTZ, P; MARSELIS,S; PATTERSON, P. L; QI, W; SILVA, C. The global ecosystem dynamics investigation: High-resolutio l ser r gi g of the Earth's forests and topography. **Science of Remote Sensing**, v. 1, p. 100002, 2020.

DUBAYAH, R; BLAIR, J. B; GOETZ, S; FATOYINBO, L; HANSEN, M; HEALEY, S; HOFTON, M; HURTT, G; KELLNER, J; LUTHCKE, S; ARMSTON, J; TANG, H; DUNCANSON, L; HANCOCK, S; JANTZ, P; MARSELIS, S; PATTERSON, P.L; QI, W; SILVA, C. THE GLOBAL ECOSYSTEM DYNAMICS INVESTIGATION: HIGH-RESOLUTION LASER RANGING OF THE EARTH'S FORESTS AND TOPOGRAPHY. **SCIENCE OF REMOTE SENSING**, v. 1, P. 100002, 2020.

DUBAYAH, R; H. TANG, J. ARMSTON, S. LUTHCKE, M. HOFTON, J. BLAIR. GEDI L2B canopy cover and vertical profile metrics data global footprint level V001. **NASA: Washington, DC, USA**, 2020.https://doi.org/10.5067/GEDI/GEDI02_B.002. (Accessed: 29 August 2021).

DUBAYAH, R; M. HOFTON, J. BLAIR, J. ARMSTON, H. TANG, S. LUTHCKE. GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002. 2021a, **distributed by NASA EOSDIS Land Processes DAAC**, https://doi.org/10.5067/GEDI/GEDI02_A.002. (Accessed 29 August 2021).

DUNCANSON, L; NEUENSCHWANDER, A; HANCOCK, S; THOMAS, N; FATOYINBO, T; SIMARD, M; ... DUBAYAH, R. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. **Remote Sensing of Environment**, v. 242, p. 111779, 2020.

DUNCANSON, L.I; NIEMANN, K.O; WULDER, M.A; 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. **Remote Sensing of Environment**, v. 114, n. 1, p. 138-154, 2010.

DURIGAN G; MELO A.C.G.D; MAX J.C.M; BOAS O.V; CONTIERI W.A; RAMOS V.S. **Manual para recuperação da vegetação de cerrado**. 3 ed. São Paulo: Secretaria do Meio Ambiente, 2011.

DURIGAN G; PILON N.A.L; ASSIS G.B; SOUZA F.M; BAITELLO J.B. **Plantas pequenas do cerrado: biodiversidade negligenciada**. 1.ed. – São Paulo: Secretaria do Meio Ambiente, 2018.

EITEN, G. Vegetação do Cerrado In: PINTO, M.N. Coord. **Cerrado: caracterização, ocupação e perspectivas**. 2.ed. Brasília: UnB/ SEMATEC, p.9-65. 1994.

EMBRAPA. **Bioma Cerrado**, 2008. Disponível em: <https://www.embrapa.br/cerrados/colecao-entomologica/bioma-cerrado>. Accessed: 19 October 2022.

ERENER, A. Remote sensing of vegetation health for reclaimed areas of Seyitömer open cast coal mine. **International Journal of Coal Geology**, v. 86, n. 1, p. 20-26, 2011.

European Centre for Medium-Range Weather Forecasts - ECMWF. **ERA 5: data documentation: Era 5**. 2020. Disponível em: <https://confluence.ecmwf.int/display/CKB/ERA5>. Accessed: 01 July 2020.

EUROPEAN SPACE AGENCY (ESA). **Sentinel-2 MSI introduction**. 2020. Disponível em: <https://sentinels.copernicus.eu/web/sentinel/userguides/sentinel-2-msi>. Accessed: 10 February 2021

EUROPEN SPACE AGENCY (ESA). **User guides - Sentinel-2 MSI – revisit and coverage**. Disponível em: https://sentinel.esa.int/web/sentinel/userguides/sentinel-2-msi/revisit-coverage. Acesso em: Accessed: 11 March 2021

EUROPEN SPACE AGENCY (ESA). **User guides - Sentinel-2 MSI -overview**. Disponível em: https://sentinel.esa.int/web/sentinel/userguides/sentinel-2-msi/overview. Accessed: 22 March 2021

FARR, T.G; ROSEN, P.A; CARO, E; CRIPPEN, R; DUREN, R; HENSLEY, S; KOBRICK, M; PALLER, M; RODRIGUEZ, E; ROTH, L. The shuttle radar topography mission. **Reviews of Geophysics**, v. 45, n. 2, 2007.

FAYAD, I; BAGHDADI, N. N; ALVARES, C. A; STAPE, J. L; BAILLY, J. S. SCOLFORO, H. F; ZRIBI, M; MAIRE, G. L. Assessment of GEDI's LiDAR data for the estimation of canopy heights and wood volume of eucalyptus plantations in Brazil. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 14, p. 7095-7110, 2021.

FEEST, A; ALDRED, T. D; JEDAMZIK, K. Biodiversity quality: a paradigm for biodiversity. **Ecological Indicators**, v. 10, n. 6, p. 1077-1082, 2010.

FELFILI, J. M. **Biogeografia do Bioma Cerrado: vegetação e solos da Chapada dos Veadeiros**. 1ª ed. Universidade de Brasília/Finatec, 2007.

FERRAZ, R.S; SOUZA, S.F; REIS, M.L.L. Laser Scanner Terrestre: teoria, aplicações e prática. **Revista Brasileira de Geomática**, v. 4, n. 2, p. 99-109, 2016.

FLORES, BM; DE SÁ DECHOUM, M; SCHMIDT, IB; HIROTA, M; ABRAHÃO, A; VERONA, L; PECORAL, L. F; CURE, M. B; GILES, A.L; BRITTO COSTA, P, PAMPLONA, M. B; MAZZOCHINI, G. G; GROENENDIJK, P; MINSKI, G. L; WOLFSDORF, G; SAMPAIO, A. B; PICCOLO, F; MELO, L; LIMA, R. F. Tropical riparian forests in danger from large savanna wildfires. **Journal of Applied Ecology**, v. 58, n. 2, p. 419-430, 2021.

FONSECA, L. D; DALAGNOL, R; MALHI, Y; RIFAI, S. W; COSTA, G. B; SILVA, T. S; ROCHA, H.R; TAVARES, I.B; BORMA, L. S. Phenology and seasonal ecosystem productivity in an Amazonian floodplain forest. **Remote Sensing**, v.11, n.13, p.1530. 2019.

FRICKER, G. A; J. A. WOLF, S. S. SAATCHI, AND T. W. GILLESPIE. Predicting spatial variations of tree species richness in tropical forests from high-resolution remote sensing. **Ecological Applications**, v. 25, n. 7, p. 1776-1789, 2015.

GAO, X; HUETE, A.R; NI, W; MIURA, T. Optical–biophysical relationships of vegetation spectra without background contamination. **Remote Sensing Of Environment**, v. 74, n. 3, p. 609-620, 2000.

GENUER, R; POGGI, J-M; TULEAU-MALOT, C. VSURF: an R package for variable selection using random forests. **The R Journal**, v. 7, n. 2, p. 19-33, 2015.

GHOSH, S.M; BEHERA, M.D. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. **Applied Geography**, v. 96, p. 29-40, 2018.

GILLESPIE, T.W, FOODY, G.M, ROCCHINI, D, GIORGI, A.P, SAATCHI, S. Measuring and modelling biodiversity from space. **Progress in Physical Geography**, v. 32, n. 2, p. 203-221, 2008.

GIONGO, M; KOEHLER, H. S; DO AMARAL MACHADO, S; KIRCHNER, F. F; MARCHETTI, M. LiDAR: princípios e aplicações florestais. **Pesquisa Florestal Brasileira**, v. 30, n. 63, p. 231-231, 2010.

GOMES, M. F; MAILLARD, P. An evaluation of RADARSAT-1 Data for the characterization of Cerrado vegetation. In: **ISPRS Annual Conference, Anais, Enschede, Holanda, CD-ROM**. 2006.

GRABSKA, E; FRANTZ, D; AND OSTAPOWICZ, K. Evaluation of machine learning algorithms for forest stand species mapping using Sentinel-2 imagery and environmental data in the Polish Carpathians. **Remote Sensing of Environment**, v. 251, p. 112103, 2020.

GURALNICK, R. P; HILL, A. W; LANE, M. Towards a collaborative global infrastructure for biodiversity assessment. **Ecology Letters**, v.8, p. 663–672. 2007.

GYAMFI-AMPADU, E; GEBRESLASIE, M; MENDOZA-PONCE, A. Evaluating multi-sensors spectral and spatial resolutions for tree species diversity prediction. **Remote Sensing**, v.13, n. 5, p. 1033. 2021.

HAKKENBERG, C. R; ATKINS, J. W; BRODIE, J. F; BURNS, P; CUSHMAN, S; JANTZ, P; KASZT, Z; QUINN, C.A; ROSE, M. D; GOETZ, S. J. Inferring alpha, beta, and gamma plant diversity across biomes with GEDI spaceborne lidar. **Environmental Research: Ecology**, v. 2, n. 3, p. 035005, 2023.

HAKKENBERG, C. R; ZHU, K; PEET, R. K; SONG, C. Mapping multi-scale vascular plant richness in a forest landscape with integrated LiDAR and hyperspectral remote-sensing. **Ecology**, v. 99, n. 2, p. 474-487, 2018.

HALL, K; REITALU, T; SYKES, M.T; PRENTICE, H.C. Spectral heterogeneity of QuickBird satellite data is related to fine-scale plant species spatial turnover in semi-natural grasslands. **Applied Vegetation Science**, v. 15, n. 1, p. 145-157, 2012.

HANCOCK, S; ARMSTON, J; HOFTON, M; SUN, X; TANG, H; DUNCANSON, L. I; KELLNER, J.R; DUBAYAH, R. The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. **Earth and Space Science**, v. 6, n. 2, p. 294-310, 2019.

HEALEY, S. P; YANG, Z; GORELICK, N; ILYUSHCHENKO, S. Highly local model calibration with a new GEDI LiDAR asset on Google Earth Engine reduces Landsat Forest height signal saturation. **Remote Sensing**, v.12, n. 17, p. 2840. 2020.

HEINZEL, J; KOCH, B. (2012). Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. **International Journal of Applied Earth Observation and Geoinformation**, v. 18, p. 101-110, 2012.

HERNÁNDEZ-STEFANONI, J. L; DUPUY, J. M; JOHNSON, K. D; BIRDSEY, R; TUN-DZUL, F; PEDUZZI, A; CAAMAL-SOSA, J. P; SANCHEZ-SANTOS, G; LÓPEZ-MERLÍN, D. Improving species diversity and biomass estimates of tropical dry forests using airborne LiDAR. **Remote Sensing**, v. 6, n. 6, p. 4741-4763, 2014.

HERNÁNDEZ-STEFANONI, J. L; GALLARDO-CRUZ, J. A; MEAVE, J. A; ROCCHINI, D; BELLO-PINEDA, J; LÓPEZ-MARTÍNEZ, J. O. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. **International Journal of Applied Earth Observation and Geoinformation**, v. 19, p. 359-368, 2012.

HERNÁNDEZ-STEFANONI, J.L; GALLARDO-CRUZ, J.A; MEAVE, J.A; ROCCHINI, D; BELLO-PINEDA, J; LÓPEZ-MARTÍNEZ, J.O. Modelling α and β-diversity in a tropical forest from remotely sensed and spatial data. **International Journal of Applied Earth Observation and Geoinformation**. v.19, p. 359–368. 2012

HOFFMANN, L; GÜNTHER, G; LI, D; STEIN, O; WU, X; GRIESSBACH, S; HENG, Y; KONOPKA, P; MÜLLER, R; VOGEL, B; AND WRIGHT, J. S. From ERA-Interim to ERA5: the considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. **Atmospheric Chemistry and Physics**, v. 19, n. 5, p. 3097-3124, 2019.

HOFTON, M; BLAIR, J.B. **Algorithm Theoretical Basis Document (ATBD) for GEDI Transmit and Receive Waveform Processing for L1 and L2 Products**. https://lpdaac. usgs.gov/documents/581/GEDI_WF_ATBD_v1.0.pdf. 2019. Accessed 13 March 2021

HUETE, A; DIDAN, K; MIURA, T; RODRIGUEZ, E. P; GAO, X; FERREIRA, L. G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1-2, p. 195-213, 2002.

HUETE, A.R. A soil-adjusted vegetation index (SAVI). **Remote sensing of environment**, v. 25, n. 3, p. 295-309, 1988.

HURLBERT, S. H. The nonconcept of species diversity: a critique and alternative parameters. **Ecology**, v. 52, n. 4, p. 577-586, 1971.

HURTT, G; ZHAO, M; SAHAJPAL, R; ARMSTRONG, A; BIRDSEY, R; CAMPBELL, E; DOLAN, K; DUBAYAH, R; FISK, J.P; FLANAGAN, S; HUANG, C; HUANG, W; JOHNSON, K; LAMB, R. MA, L; MARKS, R; O'LEARY, D; O'NEIL-DUNNE, J; SWATANTRAN, A; TANG, H. Beyond MRV: high-resolution forest carbon modeling for climate mitigation planning over Maryland, USA. **Environmental Research Letters**, v. 14, n. 4, p. 045013, 2019.

JENSEN, J. R. **Remote Sensing of the Environment. An Earth Resource Perspective**. Harlow: Person Education Limited, 2 ed. 614 p.2011.

JIN, X; WAN, L; ZHANG, Y. K; HU, G; SCHAEPMAN, M. E; CLEVERS, J. G. P. W; SU, Z. B. Quantification of spatial distribution of vegetation in the Qilian Mountain area with MODIS NDVI. **International Journal of Remote Sensing**, v. 30, n. 21, p. 5751-5766, 2009.

JORDAN, C. F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663-666, 1969.

KERSHAW Jr; J. A; DUCEY, M. J; BEERS, T. W; HUSCH, B. **Forest mensuration. Chicheste**r; Hoboken: John Wiley Sons, 5a ed; 632 p. 2017.

KHARE, S; LATIFI, H; ROSSI, S. A 15-year spatio-temporal analysis of plant β-diversity using Landsat time series derived Rao's Q index. **Ecological Indicators**, v. 121, p. 107105, 2021.

KHARE, S; LATIFI, H; ROSSI, S. Forest beta-diversity analysis by remote sensing: How scale and sensors affect the Rao's Q index. **Ecological Indicators**, v. 106, p. 105520, 2019.

KLINK, C. A; MACHADO, R. B. A conservação do Cerrado brasileiro. Belo Horizonte, **Megadiversidade**, v. 1, n. 1, p. 148-155. 2005.

KLINK, C. A; MACHADO, R. B. Conservation of the Brazilian cerrado. **Conservation Biology**, v. 19, n. 3, p. 707-713, 2005.

KÖHL, M; MAGNUSSEN, S; MARCHETTI, M. **Sampling methods, remote sensing and GIS multiresource forest inventory**, Berlin: Springer, v. 2: p. 373. 2006.

LAMB, E. G; BAYNE, E; HOLLOWAY, G; SCHIECK, J; BOUTIN, S; HERBERS, J; HAUGHLAND, D. L. Indices for monitoring biodiversity change: Are some more effective than others? **Ecological Indicators**, v. 9, n. 3, p. 432-444, 2009.

LAURIN, G. V; CHAN, J. C. W; CHEN, Q; LINDSELL, J. A; COOMES, D. A; GUERRIERO, L; FRATE, F; MIGLIETTA, F; VALENTINI, R. Biodiversity mapping in a tropical West African forest with airborne hyperspectral data. **PloS One**, v. 9, n. 6, p. e97910, 2014.

LAURIN, G. V; PULETTI, N; HAWTHORNE, W; LIESENBERG, V; CORONA, P; PAPALE, D; CHEN, QI; VALENTINI, R. Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data. **Remote Sensing of Environment**, v. 176, p. 163-176, 2016.

LEITE, R. V; SILVA, C. A; BROADBENT, E. N; DO AMARAL, C. H; LIESENBERG, V; DE ALMEIDA, D. R. A; MOHAN, M; GODINHO, S; CARDIL, A; HAMAMURA, C; FARIA, B. L; BRANCALION, P. H.S; HIRSCH, A; MARCATTI, G. E; DALLA CORTE, A. P; ZAMBRANO, A. M. A; COSTA, M, B, T; MATRICARDI, E. A.T; SILVA, A. L; GOYA, L. R.R; KLAUBERG, C. Large scale multi-layer fuel load characterization in tropical savanna using GEDI spaceborne lidar data. **Remote Sensing of Environment**, v. 268, p. 112764, 2022.

LEVIN, N; SHMIDA, A; LEVANONI, O; TAMARI, H; KARK, S. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. **Diversity and Distributions**, v. 13, n. 6, p. 692-703, 2007.

LIANG, J; CROWTHER, T. W; PICARD, N; WISER, S; ZHOU, M; ALBERTI, G; ... REICH, P. B. Positive biodiversity-productivity relationship predominant in global forests. **Science**, v. 354, n. 6309, p. aaf8957, 2016.

LIAW, A; WIENER, M. Classification and regression by randomForest. **R News**, v. 2, n. 3, p. 18-22, 2002.

LIM, K; TREITZ, P; WULDER, M; ST-ONGE, B; FLOOD, M. LiDAR remote sensing of forest structure. **Progress in Physical Geography**, v. 27, n. 1, p. 88-106, 2003.

LIU, H. Q; HUETE, A. A feedback-based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Transactions On Geoscience And Remote Sensing**, v. 33, n. 2, p. 457-465, 1995.

LIU, X; FREY, J; MUNTEANU, C; STILL, N; KOCH, B. Mapping tree species diversity in temperate montane forests using Sentinel-1 and Sentinel-2 imagery and topography data. **Remote Sensing of Environment**, v. 292, p. 113576, 2023.

LOPES, V. C; DA SILVA, R. C; DE MELO AMORIM, M. V; DE MORAES, T. P. E; BISPO, A. C; ALVAREZ, S. F; LOPES, T. S; DE SOUZA, P. B. Diversidade alfa e beta de fragmentos de Cerrado, região norte da Amazônia legal. **Observatório de la Economía Latinoamericana**, v. 21, n. 11, p. 22244-22258, 2023.

LOSCHI, R. A; PEREIRA, J. A. A; MACHADO, E. L. M; CARLOS, L; GONZAGA, A. P. D; CARMO, I. P; GOMES, D. J. S. Variações estruturais e ambientais em um contínuo de mata de galeria/cerrado stricto sensu em Itumirim, MG. **Cerne**, v.19, p. 213-227. 2013.

LOVETT, G. M; BURNS, D. A; DRISCOLL, C. T; JENKINS, J. C; MITCHELL, M. J; RUSTAD, L; SHANLEY, J. B; LIKENS, G. E; HAEUBER, R. Who needs environmental monitoring? **Frontiers in Ecology and the Environment**, v. 5, n. 5, p. 253-260, 2007.

MACHADO, C. C. C; GALVÍNCIO, J. D; DE MOURA, M. S. B; DE ARAUJO, H. F. P. Predicting plant species richness with satellite images in the largest dry forest nucleus in South America. **Journal of Arid Environments**, v. 166, p. 43-50, 2019.

MACHADO, R. B; RAMOS-NETO, M. B; PEREIRA, P. G. P; CALDAS, E. F; GONÇALVES, D. A; SANTOS, N. S; TABOR, K. STEININGER, M. **Estimativas de perda da área do Cerrado brasileiro**. Relatório Técnico Não Publicado, Conservation International do Brasil, Brasília. 2004.

MADONSELA, S; CHO, M. A; RAMOELO, A; MUTANGA, O. Remote sensing of species diversity using Landsat 8 spectral variables. **ISPRS Journal of Photogrammetry and Remote Sensing**, v.133, p. 116-127, 2017.

MADONSELA, S; CHO, M. A; RAMOELO, A; MUTANGA, O. Remote sensing of species diversity using Landsat 8 spectral variables. **ISPRS Journal of Photogrammetry and Remote Sensing,** v. 133, p. 116-127, 2017.

MAGURRAN, A. E. **Medindo a diversidade biológica**. Tradução: Vianna D.M. Curitiba: Ed. da UFPR. 261p. 2011.

MAGURRAN, A.E. **Measuring biological diversity**. Oxford: Blackwell Science, 264p. 2004.

MALLINIS, G; CHRYSAFIS, I; KORAKIS, G; PANA, E; KYRIAZOPOULOS, A. P. A random forest modelling procedure for a multi-sensor assessment of tree species diversity. **Remote Sensing**, v. 12, n. 7, p. 1210, 2020.

MAPFUMO, R. B; MURWIRA, A; MASOCHA, M; ANDRIANI, R. The relationship between satellite-derived indices and species diversity across African savanna ecosystems. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p. 306-317. 2016.

MARSELIS SM, TANG H, ARMSTON JD, CALDERS K, LABRIÈRE N, DUBAYAH R. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: A case study in Lopé National Park, Gabon. **Remote Sensing of Environment**, v. 216, p. 626-634, 2018.

MARSELIS, S. M; TANG, H; ARMSTON, J; ABERNETHY, K; ALONSO, A; BARBIER, N; BISSIENGOU, P; JEFFERY, K; KENFACK, D; LABRIERE; N; LEE. S-K; LEWIS, S.L; MEMIAGHE, H; POULSEN, J. R; WHITE, L; DUBAYAH, R. Exploring the relation between remotely sensed vertical canopy structure and tree species diversity in Gabon. **Environmental Research Letters**, v. 14, n. 9, p. 094013, 2019.

MARTINS, F. R; SANTOS, F. M. A. Técnicas usuais de estimativa da biodiversidade. **Revista Holos**, Rio claro, v.1, p. 236-267, 1999.

MELO, A. S. O que ganhamos 'confundindo' riqueza de espécies e equabilidade em um índice de diversidade? **Biota Neotropica**, v. 8, p. 21-27, 2008.

MENG, J; LI, S; WANG, W; LIU, Q; XIE, S; MA, W. Estimation of forest structural diversity using the spectral and textural information derived from SPOT-5 satellite images. **Remote Sensing**, v. 8, n. 2, p. 125, 2016.

MILHEIRAS, S. G; MACE, G. M. Assessing ecosystem service provision in a tropical region with high forest cover: Spatial overlap and the impact of land use change in Amapá, Brazil. **Ecological Indicators**, v. 99, p. 12-18, 2019.

MILLARD, K; RICHARDSON, M. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. **Remote Sensing**, v. 7, n. 7, p. 8489-8515, 2015.

MINISTÉRIO DO MEIO AMBIENTE (MMA) (1999) **Agenda 21 Brasileira – Agricultura sustentável – Produto 3 / Versão final**. Museu Emílio Goeldi / USP-

PROCAM / ATECH. (disponível em: www.fea.unicamp.br/docentes/ortega/curso/docfinal.rtf ).

MITTERMEIER R.A; TURNER W.R; LARSEN F.W; BROOKS T.M; GASCON C. Global biodiversity conservation: The critical role of hotspots. In: F.E. Zachos, Habel J.C. **Biodiversity Hotspots: distribution and protection of conservation priority areas**. Heidelberg: Springer. p. 550. 2011.

MORRIS, E.K; CARUSO, T; FISCHER, M; HANCOCK, C; OBERMAIER, E; PRATI, D; MAIER, T.S; MEINERS, T; CAROLINE, M; WUBET, T; ET AL; Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. **Ecology and Evolution**, v. 4, n. 18, p. 3514-3524, 2014

MUKUL, M; SRIVASTAVA, V; JADE, S; MUKUL, M. Uncertainties in the shuttle radar topography mission (SRTM) Heights: Insights from the indian Himalaya and Peninsula. **Scientific Reports**, v. 7, n. 1, p. 41672, 2017.

MUTOWO, G; MURWIRA, A. Relationship between remotely sensed variables and tree species diversity in savanna woodlands of Southern Africa. **International Journal of Remote Sensing**, v. 33, n. 20, p.6378-6402, 2012.

NAGENDRA, H. Opposite trends in response for the Shannon and Simpson indices of landscape diversity**. Applied Geography**, v. 22, n. 2, p. 175-186, 2002.

NERI, AV; SCHAEFER, CEG. R; SOUZA, AL; FERREIRA-JUNIOR, WG; MEIRA-NETO, JAA. Pedology and plant physiognomies in the Cerrado, Brazil. **Anais da Academia Brasileira de Ciências**, v. 85, p. 87-102, 2013.

NICHOLS, J. D; WILLIAMS, B. K. Monitoring for conservation. **Trends in Ecology Evolution**, v. 21, n. 12, p. 668-673, 2006.

NIMER, E. **Climatologia do Brasil**. Rio de Janeiro: IBGE, 1989.

NOAA, CSC. **Lidar 101: an introduction to lidar technology, data, and applications**. NOAA Coastal Services Center, 72 p. 2012.

NOSS, R. F. Indicators for monitoring biodiversity: a hierarchical approach. **Conservation Biology**, v. 4, n. 4, p. 355-364, 1990.

OGILVIE, A; POUSSIN, J. C; BADER, J. C; BAYO, F; BODIAN, A; DACOSTA, H; DIA, D; DIOP, L; MARTIN, D; SAMBOU, S. Combining multi-sensor satellite imagery to improve long-term monitoring of temporary surface water bodies in the Senegal river floodplain. **Remote Sensing**, v. 12, n. 19, p. 3157, 2020.

OLDELAND, J; WESULS, D; ROCCHINI, D; SCHMIDT, M; JÜRGENS, N. Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? **Ecological Indicators**, v. 10, n. 2, p. 390-396, 2010.

OLIVEIRA-FILHO, A. T; RATTER, J. Vegetation physiognomies and woody flora of the Cerrado biome. In: **The Cerrados of Brazil: ecology and natural history of a neotropical savanna**. Columbia University Press. p. 91-120, 2002.

PALMER, M. W; EARLS, P. G; HOAGLAND, B. W; WHITE, P. S; WOHLGEMUTH, T. Quantitative tools for perfecting species lists. **Environmetrics: The Official Journal of the International Environmetrics Society**, v. 13, n. 2, p. 121-137, 2002.

PAQUETTE, A; MESSIER, C. The effect of biodiversity on tree productivity: from temperate to boreal forests. **Global Ecology and Biogeography**, v. 20, n. 1, p. 170-180, 2011.

PEET, R.K. The measurement of species diversity. **Annual Review of Ecology and Systematics**, v. 5, n. 1, p. 285-307, 1974.

PENG, Y; FAN, M; SONG, J; CUI, T; LI, R. ASSESSMENT of plant species diversity based on hyperspectral indices at a fine scale. **Scientific Reports,** v.8, n.1, p. 1-11. 2018.

PENNINGTON, R. T; RATTER, J. **Neotropical savannas and seasonally dry forests: plant diversity, biogeography, and conservation**. CRC press, pp. 121–157. 2006.

PINHEIRO, E. D. S; DURIGAN, G. Diferenças florísticas e estruturais entre fitofisionomias do cerrado em Assis, SP, Brasil. **Revista Árvore**, v. 36, p. 181-193. 2012.

PORTILLO-QUINTERO, C. A; SÁNCHEZ-AZOFEIFA, G. A. Extent and conservation of tropical dry forests in the Americas. **Biological Conservation**, v. 143, n. 1, p. 144-155, 2010.

PORTO, AC; LINARES, JAH; NETO, GBS. Análise da estrutura e dinâmica da paisagem do Parque Nacional da Chapada dos Veadeiros. Simpósio Brasileiro de Sensoriamento Remoto-SBSR, 15º. **Anais. [...]** Curitiba, INPE, 2011, p. 3057.

QI, W; LEE, S.K; HANCOCK, S; LUTHCKE, S; TANG, H; ARMSTON, J; DUBAYAH, R; Improved Forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. **Remote Sensing of Environment**, v. 221, p. 621-634, 2019.

RAMPHERI, M; DUBE, T; DHAU, I. Use of remotely sensed data to estimate tree species diversity as an indicator of biodiversity in Blouberg Nature Reserve, South Africa. **Geocarto International**, v.37, n. 2, p. 526-542. 2022.

READ, Q. D; ZARNETSKE, P. L; RECORD, S; DAHLIN, K. M; COSTANZA, J. K; FINLEY, A. O; GRADY, J. M; HOBI, M. L; LATIMER, A. M; MALONE, S. L; OLLINGER, S. V; PAU, S; WILSON, A. M. Beyond counts and averages: Relating geodiversity to dimensions of biodiversity. **Global Ecology and Biogeography**, v. 29, n. 4, p. 696-710, 2020.

REIS, A. B. S. **Estudo De Índices De Diversidade Como Descritores De Textura Para Classificação De Imagens De Faces**. 81f. Dissertação (Mestrado em Engenharia de Eletricidade), Universidade Federal do Maranhão. São Luís, 2012.

REN, G; ZHU, A. X; WANG, W; XIAO, W; HUANG, Y; LI, G; LI, D; ZHU, J. A hierarchical approach coupled with coarse DEM information for improving the efficiency and accuracy of forest mapping over very rugged terrains. **Forest Ecology and Management**, v. 258, n. 1, p. 26-34, 2009.

REX, F. E; SILVA, C. A; DALLA CORTE, A. P; KLAUBERG, C; MOHAN, M; CARDIL, A; SILVA, V. S; ALMEIDA, D. R.A; GARCIA, M; BROADBENT, E .N; VALBUENA, R; STODDART, J;MERRICK, T; HUDAK, A. T. Comparison of statistical modelling approaches for estimating tropical forest aboveground biomass stock and reporting their changes in low-intensity logging areas using multi-temporal LiDAR data**. Remote Sensing**, v. 12, n. 9, p. 1498, 2020.

RIBEIRO, J.F; S.M. SANO E J.A. DA SILVA. 1981. Chave preliminar de identificação dos tipos fisionômicos da vegetação do Cerrado. pp. 124-133 In: XXXII Congresso Nacional de Botânica. Sociedade Botânica do Brasil, Teresina, Brasil. **Anais [...]** RIBEIRO, José Felipe; WALTER, Bruno Machado Teles. Fitofisionomias do bioma Cerrado. 1998.

RIBEIRO, J.F; WALTER, B.M.T. As principais fitofisionomias do Bioma Cerrado. In **Cerrado: Ecologia e Flora**; Sano, S.M; Almeida, S.P; Ribeiro, J.F; Eds; Embrapa Cerrados: Brasília, Brasil, pp. 151–212. 2008.

RICHARDSON, A. J; WIEGAND, C. L. Distinguishing vegetation from soil background information. **Photogrammetric engineering and remote sensing**, v. 43, n. 12, p. 1541-1552, 1977.

RICHERSON, P. J; LUM, K. L. Patterns of plant species diversity in California: relation to weather and topography. **The American Naturalist**, v. 116, n. 4, p. 504-536, 1980.

RICOTTA, C. On parametric evenness measures. **Journal of theoretical biology**, v. 222, n. 2, p. 189-197, 2003.

RIGHI, C. A; DE OLIVEIRA RISANTE, A. P; PACKER, A. P; DO COUTO, H. T. Z. Biodiversity and biomass relationships in a Cerrado stricto sensu in Southeastern Brazil. **Environmental Monitoring and Assessment**, v. 195, n. 4, p. 492, 2023.

ROBINSON, C; SAATCHI, S; CLARK, D; HURTADO ASTAIZA, J; HUBEL, A. F; GILLESPIE, T. W. Topography and three-dimensional structure can estimate tree diversity along a tropical elevational gradient in Costa Rica**. Remote Sensing**, v. 10, n. 4, p. 629, 2018.

ROCCHINI, D. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. **Remote Sensing of Environment**, v. 111, n. 4, p. 423-434, 2007.

RODRIGUEZ-GALIANO, V.F; GHIMIRE, B; ROGAN, J; CHICA-OLMO, M.J; RIGOL-SANCHEZ, P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal Of Photogrammetry And Remote Sensing**, v. 67, p. 93-104, 2012.

ROUJEAN, J. L; BREON, F. M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. **Remote Sensing of Environment**, v. 51, n. 3, p. 375-384, 1995.

ROUSE, J. W. Monitoring the vernal advancement of retrogradation of natural vegetation. **NASA/GSFC, type III, final report, greenbelt, MD**, v. 371, 1974.

SANO, E. E; FERREIRA, L. G. Monitoramento semidetalhado (escala 1:250.000) de ocupação de solos do cerrado: considerações e proposta metodológica. XII Simpósio Brasileiro de Sensoriamento Remoto. **Anais [...].** Goiânia, INPE, 16-21 abril 2005, p. 3309-3316. 2005.

SANO, E. E; ROSA, R; BRITO, J. L. S; FERREIRA, L. G. Mapeamento semidetalhado do uso da terra do Bioma Cerrado. **Pesquisa Agropecuária Brasileira**, v. 43, n. 1, p.153-156. 2008.

SANO, E. E; ROSA, R; BRITO, J. L. S; FERREIRA, L. G; **Mapeamento de cobertura vegetal do Bioma Cerrado: estratégias e resultados**. Planaltina, DF: Embrapa Cerrados, 2007.

SANTOS, J. R; NARVAES, I. S; GRAÇA, P. M. L. A; GONÇALVES, F. G. Polarimetric responses and scattering mechanisms of tropical forests in the Brazilian Amazon. In: Gary Jedlovec (NASA/MSFC-USA). Vukovar, Croatia: INTECH. **Advances on Geoscience and Remote Sensing**. v.8, p.183-206, 2009.

SANTOS, S. A. **As unidades de conservação no cerrado frente ao processo de conversão**. 2018. 105f. Dissertação (Mestrado em Geografia) - Universidade Federal de Goiás, Goiânia, 2018.

SAWYER, D. População, meio ambiente e desenvolvimento sustentável no cerrado. In: HOGAN, D. J; CARMO, R. L; CUNHA, J. M. P; BAENINGER, R. (org.). **Migração e ambiente no Centro-Oeste**. Campinas, NEPO/UNICAMP: PRONEX, p. 279-299. 2002.

SCHÄFER, E; HEISKANEN, J; HEIKINHEIMO, V; PELLIKKA, P. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. **Ecological Indicators**, v. 64, p. 49-58, 2016.

SCHUSTER, C; FÖRSTER, M; KLEINSCHMIT, B. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. **International Journal of Remote Sensing**, v. 33, n. 17, p. 5583-5599, 2012.

SELLERS, P. J. Canopy reflectance, photosynthesis and transpiration. **International Journal of Remote Sensing**, v. 6, n. 8, p. 1335-1372, 1985.

SHANNON C. E; WEAVER W. **The mathematical theory of communication**. University of Illinois Press, 1963.

SHOKO C, MUTANGA O, DUBE T. Remotely sensed C3 and C4 grass species aboveground biomass variability in response to seasonal climate and topography. **African Journal of Ecology**, v. 57, n. 4, p. 477-489, 2019.

SILVA B, ALAVA-N UNEZ P, STROBL S, BECK E, BENDIX J. Area-wide evapotranspiration monitoring at the crown level of a tropical mountain rain forest. **Remote Sensing of Environment**, v. 194, p. 219-229, 2017.

SILVA, C.A; DUNCANSON, L; HANCOCK, S; NEUENSCHWANDER, A; THOMAS, N; HOFTON, M; FATOYINBO, L; SIMARD, M; MARSHAK, C.Z; ARMSTON, J; LUTCHKE, S; DUBAYAH, R; Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. **Remote Sensing of Environment**, v. 253, p. 112234, 2021.

SIMONSON, W. D; ALLEN, H. D; COOMES, D. A. Use of an airborne lidar system to model plant species composition and diversity of Mediterranean oak forests. **Conservation Biology**, v. 26, n. 5, p. 840-850, 2012.

SIMPSON, E. H. Measurement of diversity. **Nature**, v. 163, n. 4148, p. 688-688, 1949

ŠKERLAK, B; SPRENGER, M; WERNLI, H. A global climatology of stratosphere–troposphere exchange using the ERA-Interim data set from 1979 to 2011**.** **Atmospheric Chemistry and Physics**, v. 14, n. 2, p. 913-937, 2014.

SOUZA, C.M; SHIMBO, J.Z; ROSA, M.R; PARENTE, L.L; ALENCAR, A.A; RUDORFF, B.F.T; HASENACK, H; MATSUMOTO, M; FERREIRA, L.G; SOUZA-FILHO, P.W.M; DE OLIVEIRA, S.W; ROCHA, W.F; FONSECA, A.V; MARQUES, C.B; DINIZ, C.G; COSTA, D; MONTEIRO, D; ROSA, E.R; V´ELEZ-MARTIN, E; WEBER, E.J; LENTI, F.E.B; PATERNOST, F.F; PAREYN, F.G. C; SIQUEIRA, J.V; VIERA, J.L; NETO, L.C.F; SARAIVA, M.M; SALES, M.H; SALGADO, M.P. G; VASCONCELOS, R; GALANO, S; MESQUITA, V.V; AZEVEDO, T. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and earth engine. **Remote Sensing**, v. 12, n. 17, p. 2735, 2020.

TORBICK, N; LEDOUX, L; SALAS, W; ZHAO, M. Regional mapping of plantation extent using multisensor imagery. **Remote Sensing**, v. 8, n. 3, p. 236, 2016.

TORRESANI, M; ROCCHINI, D; SONNENSCHEIN, R; ZEBISCH, M; MARCANTONIO, M; RICOTTA, C; TONON, G. Estimating tree species diversity from space in an alpine conifer forest: The Rao's Q diversity index meets the spectral variation hypothesis. **Ecological Informatics**, v. 52, p. 26-34, 2019.

TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, v. 8, n. 2, p. 127-150, 1979.

VAIPHASA, C; SKIDMORE, A.K; DE BOER, W.F; VAIPHASA, T. A hyperspectral band selector for plant species discrimination. **ISPRS Journal of Photogrammetry and Remote Sensing,** v. 62, n. 3, p. 225-235, 2007.

VALBUENA, R; O'CONNOR, B; ZELLWEGER, F; SIMONSON, W; VIHERVAARA, P; MALTAMO, M; SILVA, C. A; ALMEIDA, D. R.A; DANKS, F; MORSDORF, F; CHIRICI, G; LUCAS, R; COOMES, D. A; COOPS, N. C. Standardizing ecosystem morphological traits from 3D information sources. **Trends in Ecology Evolution**, v. 35, n. 8, p. 656-667, 2020.

WEVER, C; LINDENBERGER, J. Experiences of 10 years laser scanning. In: **Photogrammetric Week 99**. P. 125-132.1999.

YANG, L; MENG, X; ZHANG, X. SRTM DEM and its application advances. **International Journal of Remote Sensing**, v. 32, n. 14, p. 3875-3896, 2011.

YOCCOZ, N. G; NICHOLS, J. D; BOULINIER, T. Monitoring of biological diversity in space and time. **Trends in Ecology Evolution**, v. 16, n. 8, p. 446-453, 2001.

ZELLWEGER, F; A. BALTENSWEILER, C. GINZLER, T. ROTH, V. BRAUNISCH, H. BUGMANN, AND K. BOLLMANN. Environmental predictors of species richness in forest landscapes: abiotic factors versus vegetation structure. **Journal of Biogeography**, v. 43, n. 6, p. 1080-1090, 2016.