Universidade do Estado de Mato Grosso
Campus de Nova Xavantina
Programa de Pós-Graduação em Ecologia e Conservação

Relação entre floresta e a agricultura: como a vegetação nativa da paisagem influencia a produtividade agrícola?

Bianca Ferraz Rebelatto

Nova Xavantina-MT
2023

Universidade do Estado de Mato Grosso
Campus de Nova Xavantina
Programa de Pós-Graduação em Ecologia e Conservação

Relação entre floresta e a agricultura: como a vegetação nativa da paisagem influencia a produtividade agrícola?

Bianca Ferraz Rebelatto

Dissertação apresentada à Coordenação do Programa de Pós-graduação em Ecologia e Conservação da Universidade do Estado de Mato Grosso, *Campus* de Nova Xavantina, como parte dos requisitos para obtenção do título de Mestre em Ecologia e Conservação.

Orientador: Prof. Dr. Divino Vicente Silvério
Coorientadora: Dra. Ludmila Maria Rattis Teixeira

Nova Xavantina-MT
2023

Ficha catalográfica elaborada pelo Setor de Processamento Técnico da Divisão de Biblioteca da UNEMAT Catalogação de Publicação na Fonte. UNEMAT - Unidade padrão

Rebelatto, Bianca Ferraz.
Relação entre floresta e a agricultura: como a vegetação nativa da paisagem influencia a produtividade agrícola / Bianca Ferraz Rebelatto. - Nova Xavantina, 2023.
49f.: il.

Universidade do Estado de Mato Grosso, Ecologia e Conservação, Nova Xavantina.
Orientador: Divino Vicente Silvério.
Coorientador: Ludmila Maria Rattis Teixeira.

1. Sensoriamento remoto. 2. Modelagem. 3. Ecologia de paisagem. I. Silvério, Divino Vicente. II. Teixeira, Ludmila Maria Rattis. III. Título.

UNEMAT / MTSCB CDU 621.38

AGRADECIMENTOS

Inicio com meus profundos agradecimentos ao meu orientador, Divino Vicente Silvério, e a minha coorientadora, Ludmila Rattis, pelos ensinamentos, ideias, apoio e orientações ao longo dessa jornada. Aprendi muito com vocês, e levarei essa nova bagagem de conhecimento para a vida toda. Muito obrigada!

Agradeço aos meus amigos, Carla Heloísa e Jairo Matos, sempre dispostos a uma ligação aleatória e a encontros quando conseguimos estar na mesma cidade. Sempre as melhores cervejadas e conversas sobre ciência e a vida que tornaram tudo mais tranquilo nesse processo.

Agradeço imensamente aos amigos que fiz durante o mestrado: Carmino e Lorrayne. Obrigada pelas fofocas, cafezinhos, troca de ideias, lágrimas, gargalhadas, e à ajuda imensurável no meu trabalho. Vocês são 1000!!! E não poderia deixar de citar Marcos, Carliane, Nando, Mateus e Josi, que ajudaram a tornar o final desse processo mais leve e feliz.

Faço meu agradecimento especial ao meu grande amigo de longa data, Eduardo Marques, pela troca de ideias, momentos de descontração e apoio em partes desse processo que foi extremamente importante.

Agradeço aos meus pais, Klenia e Gilson, por estarem presentes mesmo à distância.

Agradeço ao professor Paulo Brando pela disposição em conversar sobre o projeto e por sua participação na construção das ideias.

Ao Laboratório de Ecologia de Transições da UNEMAT, em especial ao Professor Eddie Lenza, pelo espaço cedido no LECOT e pela disposição nas trocas de ideias e correção do projeto. Agradeço também ao Laboratório de Geotecnologias e Produção de Softwares da UFRA, supervisionado também pelo prof. Divino, pelo espaço cedido e a todos os integrantes (Francimary, Camile, Rafael, Amanda, Ronaldo, Vitoria, Alycia) pelas boas conversas ao longo dos dias.

Ao Instituto Federal de Tecnologia de Zurique (ETHZ) e ao Centro Helmholtz de Pesquisa Ambiental (UFZ) pelo suporte científico e técnico especializado. A Sonia I. Seneviratne, Jakob Zscheischler e equipe pelo suporte e discussões necessárias para o desenvolvimento dessa pesquisa, além da contribuição para o meu crescimento como estudante pesquisadora. Agradeço aos colegas e amigos Ryan, Fulden, Felix, Ahmed, Peter, Mohit, Lily, Sifang, Shijie e Emanuelle que tive a oportunidade de conhecer. E meu agradecimento especial a Andreia Ribeiro que esteve presente em todas as etapas estatísticas do primeiro capítulo desse trabalho e, juntamente o com o André, esteve presente nas etapas incríveis dos dois capítulos do meu intercâmbio — obrigado por compartilharem suas experiências e amizades.

Agradeço à Universidade do Estado de Mato Grosso e ao Programa de Pós-Graduação em Ecologia e Conservação de Nova Xavantina-MT pela estrutura e suporte durante o mestrado. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo por todo o período do mestrado e pela bolsa de estudo no exterior. Ao Instituto de Pesquisa Ambiental da Amazônia (IPAM) pelo suporte necessário para o desenvolvimento desse projeto.

Sumário

LISTA DE FIGURAS 7
LISTA DE TABELAS 9
RESUMO 10
ABSTRACT 11
1. Introdução 12
2. Material e Métodos 14
2.1 Área de estudo 14
2.2 Modelo de produtividade agrícola 16
2.2.1 Dados utilizados 17
2.2.2 Análise de sensoriamento remoto 17
2.2.3 Análise estatística 19
2.3 Relação entre a produtividade agrícola e a vegetação nativa da paisagem 22
2.3.1 Dados utilizados 23
2.3.2 Análise de sensoriamento remoto 23
2.3.3 Análise estatística 25
3. Resultados 26
3.1 Modelo de produtividade agrícola 26
3.2 Importância da vegetação nativa para a produtividade agrícola 29
4 Discussão 32
4.1 Modelo de produtividade agrícola: escolha dos preditores e performance do modelo 32
4.2 Importância da vegetação para a produtividade agrícola 33
5. Conclusão 35
6. Referências bibliográficas 36
7. Anexo 40

LISTA DE FIGURAS

Figura 1. Localização de 63 propriedades rurais na região do MATOPIBA (porção dos estados do Maranhão, Tocantins, Piauí e Bahia) e na região Nordeste Matogrossense, abrangendo os biomas Cerrado e Amazônia. Na área de linha contínua, os pontos representam as propriedades rurais com dados disponibilizados pelo Instituto de Pesquisa Ambiental da Amazônia (IPAM), e os polígonos de cor roxa são as áreas de soja adotadas para essas propriedades. Na área de linha pontilhada, os polígonos em laranja são áreas de soja adotadas para a extrapolação espacial do modelo de produtividade agrícola. Ambas as áreas de soja são de sequeiro com confirmação temporal e espacial de quatro a oito anos de permanência de cultivo dentro das safras de 2009/10 a 2016/17, e foram obtidas pela classificação disponibilizada por Spera et al. (2016). 15

Figura 2. Esquema da primeira etapa do trabalho de criação do modelo de produtividade agrícola e do mapa de produtividade gerado por extrapolação espacial. 1 - dados observados de entrada (em amarelo) e variáveis espaciais (em azul), seleção dos melhores preditores e validações para criação do modelo final; 2 - extrapolação espacial do modelo utilizando a classificação de cobertura do solo de Spera et al. (2016) para a construção do mapa de produtividade agrícola regional predita. 16

Figura 3. Esquema da segunda etapa do trabalho de verificação da importância da vegetação nativa para a produtividade agrícola de soja. 1 – Construção do mapa de vegetação pelos dados de cobertura do solo do MapBiomas (2022) para toda a área do mapa de produtividade agrícola predito; 2 – criação de 100 pontos aleatórios repetidos 20 vezes (2000 pontos), e de diferentes escalas de efeito em *buffers* de raios de 10km a 50km; 3 – amostragem da média de soja em 10km de *buffer* e da vegetação para todas as escalas e análises estatísticas. 22

Figura 4. Ilustração de um dos grupos de randomização com 100 pontos aleatórios de amostragem sobre a área de extrapolação de produtividade de soja de sequeiro predita. Ao redor da área de soja estão as classes de floresta e outras formações vegetais para o ano de 2016 (MapBiomas, 2022). Para cada ponto foram gerados *buffers* com tamanho máximo de distância de 10 a 50 km (variando de 5 em 5 km) de raio em relação ao ponto central (esquema ao lado direito). 24

Figura 5. Gráfico de seleção de preditores pelo maior coeficiente de determinação ajustado (R^2 ajustado) obtido pelo método de seleção de *stepwise regression*. 27

Figura 6. *Forest plot* dos coeficientes (valores sobre os pontos), erro padrão (comprimento das retas), e nível de significância (se **, $p \leq 0,05$; se ***, $p \leq 0,01$) dos preditores do modelo misto de produtividade agrícola (LMM). Valores previamente padronizados. 28

Figura 7. Comparação entre os dados de produtividade agrícola predito e observado (IBGE) a nível municipal utilizando *boxplot* (**A**) e gráfico um-pra-um da produtividade predita pela produtividade observada (IBGE), em que cada ponto é um município (**B**). 28

Figura 8. Coeficiente de determinação (R^2) da regressão entre soja predita e a proporção de floresta presente na paisagem. A soja predita foi amostrada pela sua média para *buffers* de 10 km de raio, e a vegetação pela sua proporção em *buffers* variando entre 10 e 50 km, sendo as amostras anuais para uma série temporal entre 2010 e 2017. Cada ponto representa a média de 20 grupos contendo 100 pontos gerados aleatoriamente sobre pixels de soja de sequeiro. 30

Figura 9. Produtividade de soja predita (t/ha) explicada (R^2) pela proporção de área coberta por vegetação (%) (A) e pela proporção de área coberta por floresta (%) (B) presente na paisagem agrícola em um raio de 30 km de buffer.....31
Figura 11. Correlações para a média anual de produtividade de soja predita com as métricas obtidas pela regressão linear segmentada. *Slope* 1, reta que antecede o ponto de quebra (A). *Slope* 2, reta que procede o ponto de quebra (B). Ponto de quebra, porcentagem de importância máxima de proporção de floresta dentro de um buffer de 30 km de raio para a produtividade agrícola (C).....32

LISTA DE TABELAS

Tabela 1. Descrição dos potenciais preditores para o modelo de produtividade agrícola. Para cada variável foi considerado um mês da estação de crescimento da soja (outubro a março). 18
Tabela 2. Equações do viés, raiz do erro quadrado médio (RMSE) e erro absoluto médio (MAE), com o ŷ representando o valor da produtividade de soja predita, y o valor das observações e N o número amostral. 21
Tabela 3. Médias anuais e geral de produtividade de soja de sequeiro predita e observada (IBGE) calculadas a nível municipal para os anos de colheita entre 2010 e 2017, e valor de viés. 29
Tabela 4. Sumário dos modelos lineares mistos (LMM) utilizados para avaliar a precipitação diária média como variável aleatória nos modelos que verificam o efeito da vegetação nativa e da floresta sobre a produtividade de soja predita. 31

RESUMO

A remoção em grande escala da vegetação nativa nos trópicos impacta diretamente vários serviços ecossistêmicos como, por exemplo, a produção de chuvas regulares e o equilíbrio climático. Assim, o avanço do desmatamento e seus efeitos para o clima podem potencialmente reduzir as áreas favoráveis ao cultivo agrícola no Brasil e, com isso, a produtividade final das lavouras. Os principais objetivos deste trabalho foram: a) criar um modelo estatístico/empírico de produtividade agrícola da soja com base nas variáveis climáticas e de refletância dos campos cultivados para a região do MATOPIBA, Mato Grosso e Goiás; b) avaliar se a vegetação nativa da paisagem contribui para a produtividade de soja na mesma região. Para criar o modelo de produtividade agrícola, nós utilizamos modelos lineares mistos (LMM) incluindo como variável resposta a produtividade de soja de 63 propriedades rurais (safras 2009/10 e 2016/17) e como preditores variáveis climáticas e o índice de vegetação melhorado (EVI). Em seguida, avaliamos a qualidade do modelo por meio de comparações com os dados de produtividade agrícola municipal do IBGE. Para avaliar a contribuição da vegetação nativa para a produtividade de soja, utilizamos o modelo para criar um mapa de produtividade de soja, e em seguida, selecionamos 2000 pontos aleatorizados para os quais calculamos a proporção de vegetação nativa na paisagem em um *buffer* de 30 km. O melhor modelo de produtividade da soja incluiu como preditores o Índice de Vegetação Melhorado (EVI), a Temperatura Diurna da Superfície da Terra (DLST), o Déficit de Pressão de Vapor (VPD) e o Número de Dias Secos (NDD). O coeficiente de determinação condicional (R^2c) de 71%, com relação positiva do rendimento da soja com o EVI e negativa coma DLST, VPD e NDD. A proporção da vegetação nativa na paisagem foi positivamente correlacionada com a produtividade de soja em todos os anos da série temporal, com R^2 de até 50%. A regressão segmentada apresentou maior *slope* no primeiro segmento, com ponto de quebra máximo de 26% da vegetação nativa, indicando forte redução na produtividade em paisagens com cobertura de vegetação menores. O *slope* da regressão no segundo segmento foi positivo para a maioria dos anos avaliados, indicando que o aumento na quantidade de vegetação nativa da paisagem resulta em maior produtividade de soja. Nossos resultados indicam que a floresta contribui para a produtividade da cultura da soja em função dos benefícios dos serviços ecossistêmicos locais. Os resultados deste trabalho permitirão a elaboração de mapas de produtividade regional e a promoção de práticas ambientais sustentáveis para a produção de alimentos que protegem os múltiplos serviços ecossistêmicos que a floresta fornece.

ABSTRACT

The large-scale removal of native vegetation in the tropics directly impacts several ecosystem services such as, for example, the production of regular rainfall and climate balance. Thus, the advance of deforestation and its effects on the climate can potentially reduce the areas favorable to agricultural cultivation in Brazil and, with that, the final productivity of the crops. The main objectives of this work were: a) to create a statistical/empirical model of soybean agricultural productivity based on climatic and reflectance variables of cultivated fields for the MATOPIBA region, Mato Grosso and Goiás; b) evaluate whether the landscape's native vegetation contributes to soybean productivity in the same region. To create the agricultural productivity model, we used mixed linear models (LMM) including as a response variable the soybean productivity of 63 rural properties (2009/10 and 2016/17 harvests) and as predictors climate variables and the improved vegetation index (EVI). We then evaluated the model's quality through comparisons with IBGE's municipal agricultural productivity data. To assess the contribution of native vegetation to soybean productivity, we used the model to create a soybean productivity map, and then selected 2000 randomized points for which we calculated the proportion of native vegetation in the landscape in a 30 km buffer. The best soybean productivity model included as predictors the Improved Vegetation Index (EVI), the Daytime Land Surface Temperature (DLST), the Vapor Pressure Deficit (VPD) and the Number of Dry Days (NDD). The conditional determination coefficient (R^2c) of 71%, with a positive relationship between soybean yield and EVI and a negative relationship with DLST, VPD and NDD. The proportion of native vegetation in the landscape was positively correlated with soybean productivity in all years of the time series, with R^2 of up to 50%. The segmented regression showed a higher slope in the first segment, with a maximum breakpoint of 26% of the native vegetation, indicating a strong reduction in productivity in landscapes with smaller vegetation cover. The regression slope in the second segment was positive for most of the evaluated years, indicating that the increase in the amount of native vegetation in the landscape results in higher soybean productivity. Our results indicate that the forest contributes to the productivity of the soybean crop due to the benefits of local ecosystem services. The results of this work will allow the elaboration of regional productivity maps and the promotion of sustainable environmental practices for food production that protect the multiple ecosystem services that the forest provides.

1. Introdução

A expansão agropecuária sobre a vegetação nativa nos biomas Cerrado e Amazônia tem acarretado transformações significativas nas paisagens, sendo a principal força motriz do desmatamento na região, com mais de 26 Mha e 44 Mha convertidos entre 1985 e 2020 em ambos os biomas, respectivamente (MapBiomas, 2021; Morton et al., 2006; Nogueira et al., 2008, 2007). Apenas os estabelecimentos agropecuários representavam 41,3% do território brasileiro em 2017 (IBGE Censo Agropecuário, 2017), e a tendência para as lavouras de grãos é um aumento 19,1% de área até o ano de 2032 (MAPA, 2023).

Importantes regiões de crescimento agrícola são os estados do Mato Grosso, Goiás e a região do Matopiba composta pelos estados do Maranhão, Tocantins, Piauí e Bahia (Miranda et al., 2014). Juntamente a essa conversão de área cresce também a produção agrícola (Gasques et al., 2018), quando somadas estas regiões no ano de 2020, contribuíam com aproximadamente 30% no Valor Adicionado Bruto (VAB) da agropecuária (IBGE, 2023). No entanto, a conversão de uma vegetação predominantemente arbórea em lavouras pode alterar condições do ambiente que são importantes para a própria produção.

A diminuição da cobertura vegetal nativa afeta diretamente serviços ecossistêmicos produzidos pelas florestas, como, por exemplo, o armazenamento de carbono, a regulação climática e o equilíbrio hidrológico (Davidson et al., 2012; Leite-Filho et al., 2021; Rodrigues et al., 2022; Silvério et al., 2015). Assim, se por um lado o desmatamento fornece mais áreas para o cultivo agrícola, por outro, ele pode afetar o equilíbrio dos ecossistemas e a produção de serviços que são essenciais para uma agricultura mais sustentável.

A produção agrícola das grandes *commodities* no Brasil depende fortemente de um regime de precipitação estável. Cerca de 90% das áreas cultivadas do país não são irrigadas (ANA, 2021) e depende de chuvas regulares para sua produção. Grande parte das áreas de agricultura dependem da ciclagem hidrológica e de redistribuição de umidade para a regularidade da precipitação, processos vulneráveis ao desmatamento em larga escala, como já documentado para algumas regiões (Leite-Filho et al., 2019). O desmatamento de partes do Mato Grosso e Rondônia foi o principal fator associado a um atraso no início da estação chuvosa de cerca de três semanas, reduzindo assim a janela temporal onde o cultivo agrícola é possível. Adicionalmente, o aumento da temperatura média global, decorrente das mudanças climáticas, pode reduzir as áreas propícias para o cultivo agrícola no Brasil (Rattis et al., 2021). Assim, culturas menos tolerantes a altas temperaturas são especialmente afetadas pelo aumento do

período de exposição a temperaturas acima de seus limites fisiológicos, resultando em perda de atividade enzimática e considerável redução na produtividade (Bieto e Talon, 1996).

Além da estabilização do clima em escala regional, a vegetação nativa na paisagem local pode trazer diversos outros benefícios para a produção agrícola. Diversos serviços ecossistêmicos fornecidos pelos ecossistemas naturais beneficiam a produção de alimentos, incluindo a polinização (Garibaldi et al., 2013; Monasterolo et al., 2015), o controle de pragas agrícolas por inimigos naturais (Mori et al., 2017), e a moderação das condições climáticas extremas nas lavouras (Chen et al., 1999; Flach et al., 2021). No entanto, é necessário compreender a escala de efeito e a magnitude destes benefícios fornecidos pela vegetação natural para produtividade agrícola do país.

Em relação à cultura da soja, existem evidências que apontam para a possibilidade de redução na produtividade devido ao aumento do desmatamento em larga escala e à perda de serviços ecossistêmicos em escala local. Estudos de modelagem já demonstraram perdas passadas e futuras na produtividade da soja decorrentes da diminuição dos serviços ecossistêmicos (Flach et al., 2021; Oliveira et al., 2013). Essas evidências indicam que a expansão da agricultura sobre áreas florestais e de savana terá um impacto negativo na produtividade das lavouras, acarretando perdas tanto para os remanescentes de vegetação natural quanto para o setor agrícola. Diante desses desafios, compreender a relação entre a redução da produtividade da soja e os serviços ecossistêmicos fornecidos pela paisagem local é crucial para uma gestão mais sustentável e eficaz da agricultura.

A combinação de diferentes técnicas, como sensoriamento remoto e modelagem estatística, permite uma análise mais aprofundada e abrangente dos efeitos do desmatamento na produtividade agrícola. Com a disponibilidade de dados de sensoriamento remoto de alta resolução e o uso de modelos matemáticos, combinados com dados de produtividade coletados em campo, temos uma abordagem científica que nos permite identificar possíveis padrões de saúde das culturas agrícolas, como soja e milho, e avaliar essa condição em relação à proximidade da vegetação nativa. Os modelos matemáticos criados com base em valores reais devem ser testados com dados de campo que sejam independentes e conhecidos. Apesar disso, esta ferramenta é extremamente útil para compreender o comportamento das culturas agrícolas em relação às condições edafoclimáticas e biofísicas em diferentes escalas espaciais e temporais.

Neste estudo, analisamos o efeito da vegetação nativa total e, também, especificamente, das florestas sobre a produtividade agrícola da soja. Para isso, criamos um modelo de

produtividade agrícola com dados de campo e de sensoriamento remoto, com índice de vegetação e informações climáticas. Em seguida, usamos os resultados preditivos deste modelo para avaliar possíveis contribuições da vegetação nativa da paisagem para a produtividade agrícola de soja. Essa abordagem nos permite compreender melhor a influência da floresta na produtividade agrícola para desenvolver estratégias de manejo sustentável e conservar os serviços ecossistêmicos fornecidos pela vegetação natural.

2. Material e Métodos

2.1 Área de estudo

Este trabalho teve como área de estudo parte da região do MATOPIBA e os estados de Mato Grosso, Goiás e Distrito Federal (Figura 1). A região abrange principalmente o bioma Cerrado, em que os solos, apesar de bem desenvolvidos, possuem baixa fertilidade e alta acidez, com pH variando entre 4,3 e 5,5 (Yamada, 2005). Toda a região é classificada como uma zona tropical com temperaturas elevadas do tipo Aw, com chuva no verão e seca no inverno, e do tipo Am, um clima tropical úmido e de monção (Alvares et al., 2013). Para toda a nossa região, a precipitação média diária varia entre 1,7 e 6,4 mm (Funk et al., 2015), e a temperatura média da superfície entre 25,3 e 39,0 ºC (Wan et al., 2021) (Anexo 1).

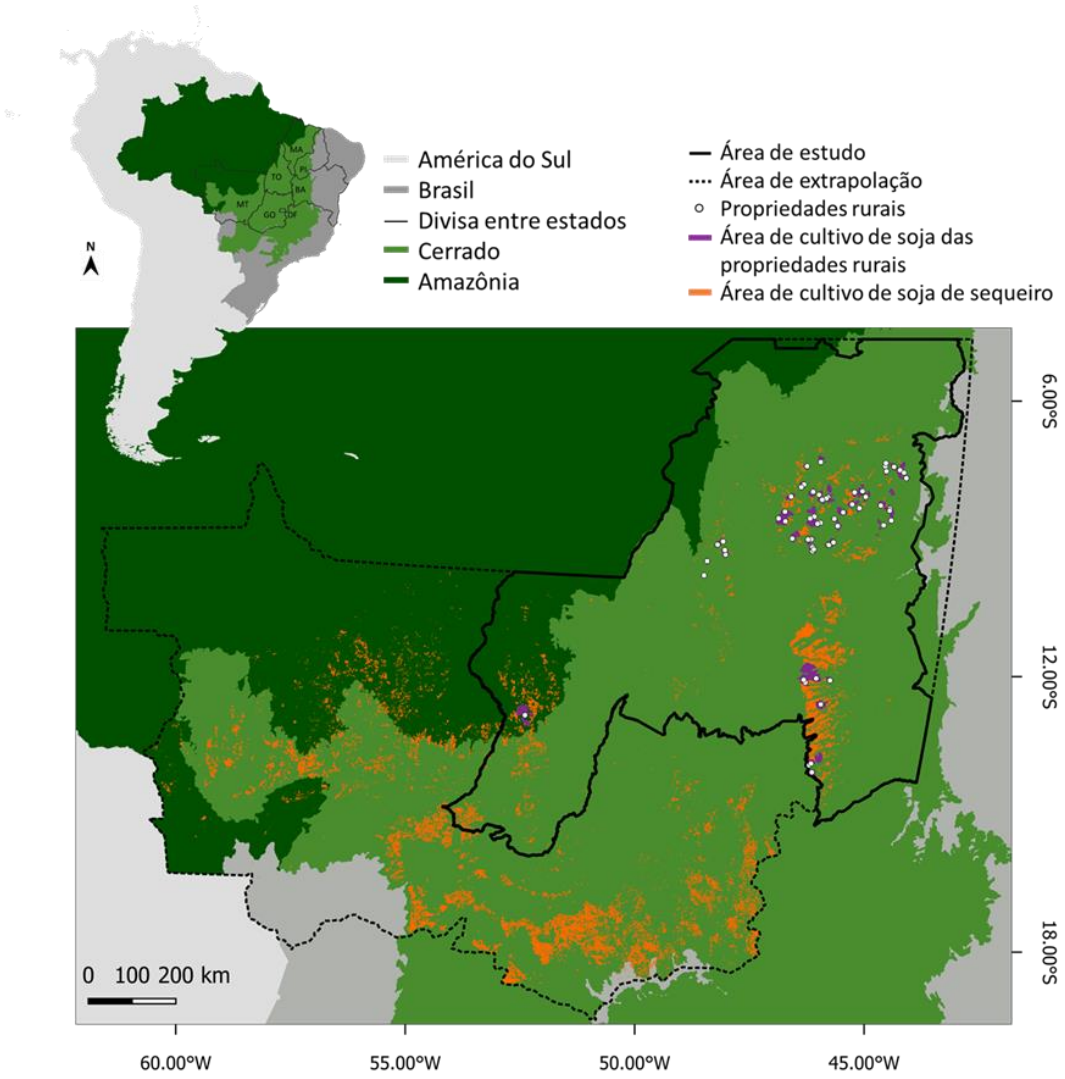


Figura 1. Localização de 63 propriedades rurais na região do MATOPIBA (porção dos estados do Maranhão, Tocantins, Piauí e Bahia) e na região Nordeste Matogrossense, abrangendo os biomas Cerrado e Amazônia. Na área de linha contínua, os pontos representam as propriedades rurais com dados disponibilizados pelo Instituto de Pesquisa Ambiental da Amazônia (IPAM), e os polígonos de cor roxa são as áreas de soja adotadas para essas propriedades. Na área de linha pontilhada, os polígonos em laranja são áreas de soja adotadas para a extrapolação espacial do modelo de produtividade agrícola. Ambas as áreas de soja são de sequeiro com confirmação temporal e espacial de quatro a oito anos de permanência de cultivo dentro das safras de 2009/10 a 2016/17, e foram obtidas pela classificação disponibilizada por Spera et al. (2016).

2.2 Modelo de produtividade agrícola

Para elaborar o modelo de produtividade da soja para a área de estudo, utilizamos uma regressão linear múltipla da produtividade observada de propriedades rurais da região em função de preditores associados a reflectância da vegetação nas áreas de cultivo e variáveis de clima (Figura 2). Mais especificamente, seguimos os seguintes passos (Anexo 2): a) utilizamos como variável resposta os dados da produtividade de soja de 63 propriedades rurais da região do MATOPIBA; b) utilizamos produtos de sensoriamento remoto para obter o Índice de Vegetação Melhorado (EVI) para os polígonos com o cultivo de soja na propriedade, e também outras cinco variáveis de clima (temperatura da superfície, números de dias sem chuva, déficit de pressão de vapor, *dry spell* e déficit hídrico climático) que compuseram as nossas variáveis explicativas; c) todas estas variáveis foram geradas mensalmente para a janela de cultivo (outubro a março; Anexo 5); d) para gerar o modelo final e avaliar a importância de cada preditor em cada mês no modelo regressivo, utilizamos um procedimento de seleção de modelos baseado no coeficiente de determinação ajustado ($R^2_{aj.}$) e consideramos uma variável aleatória, compondo um modelo de regressão linear misto; e) o melhor modelo foi utilizado para gerar um mapa de produtividade regional para todas as áreas com cultivo de soja de sequeiro. A seguir, detalhamos cada uma destas etapas.

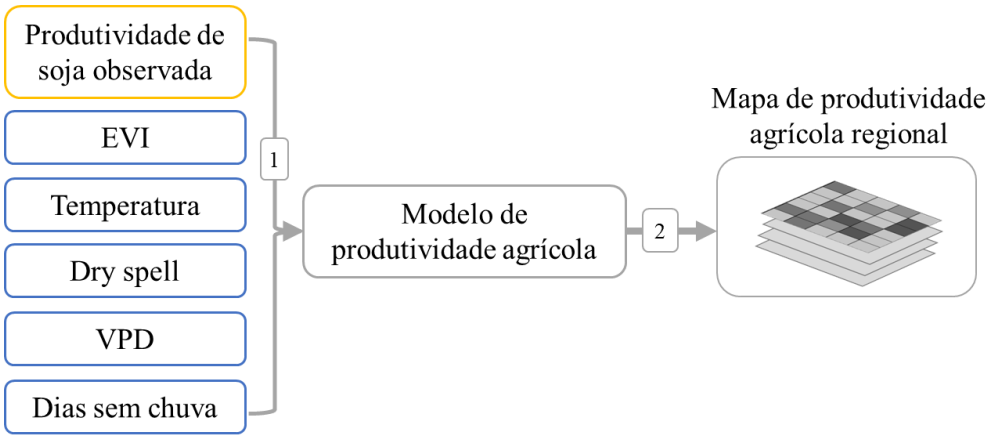


Figura 2. Esquema da primeira etapa do trabalho de criação do modelo de produtividade agrícola e do mapa de produtividade gerado por extrapolação espacial. 1 - dados observados de entrada (em amarelo) e variáveis espaciais (em azul), seleção dos melhores preditores e validações para criação do modelo final; 2 - extrapolação espacial do modelo utilizando a classificação de cobertura do solo de Spera et al. (2016) para a construção do mapa de produtividade agrícola regional predita.

2.2.1 Dados utilizados

2.2.1.1 Produtividade de soja

Os dados de produtividade de soja estão concentrados na região do MATOPIBA, uma das fronteiras agrícolas brasileiras composta pelos estados do Maranhão, Tocantins, Piauí e Bahia, e no Mato Grosso (Figura 1). Utilizamos dados da produtividade de soja (*Glycine max* (L.) Merr.) de 63 propriedades rurais (Anexo 2), incluindo os seus pontos de georreferenciamento, que foram disponibilizados pelo Instituto de Pesquisa Ambiental da Amazônia (IPAM). As propriedades têm como cultivo principal a soja de sequeiro (áreas agrícolas não irrigadas), com dados de produtividade agrícola em uma série temporal de oito anos a partir de 2010 até 2017 (safra 2009/10 à safra 2016/17; Anexo 4 e 5).

Comparamos os dados das 63 fazendas com os dados de produtividade de soja a nível municipal do Instituto Brasileiro de Geografia e Estatística (IBGE) (Anexo 6A), em que utilizamos as médias dos respectivos anos e municípios das propriedades rurais. Posteriormente, realizamos uma análise de tendência linear sobre a série temporal de produtividade (Anexo 6B) a fim de excluir possíveis efeitos não climáticos, como a tecnificação agrícola, que poderiam mascarar os efeitos da vegetação nas análises posteriores (Páscoa et al., 2017; Ribeiro et al., 2019).

2.2.2 Análise de sensoriamento remoto

2.2.2.1 Área de soja

Utilizamos um mapeamento das áreas de cultivo agrícola para relacionar as coordenadas das 63 propriedades aos polígonos de soja sendo cultivados. Para isso, lançamos mão do produto de mapeamento disponibilizado por Spera et al. (2016) para identificação dos polígonos das áreas de soja de sequeiro para cada propriedade.

O mapeamento de agricultura produzido por Spera et al. (2016) abrange os estados do Mato Grosso, Goiás, Distrito Federal e Tocantins, e uma porção dos estados do Maranhão, Piauí e Bahia, com uma resolução de 250 m (MODIS) de 2001 até 2017. Este produto identifica áreas de soja de apenas cultivo único e de soja de cultivo duplo com milho e algodão separando por áreas de irrigação, o que permitiu que selecionássemos apenas a soja considerada de sequeiro, assim como é o caso dos dados das nossas propriedades.

Para obter nosso produto da área de soja, fizemos uma sobreposição dos pixels desse produto dentro de cada propriedade para a série temporal de 2009 a 2017. Em seguida, filtramos os polígonos de soja que foram cultivadas por períodos maiores que quatro anos consecutivos, assegurando maior segurança na associação entre os dados reportados pelo produtor e os polígonos de soja efetivamente cultivados à época (Figura 1). Esse e os demais processos descritos a seguir foram realizados utilizando o programa QGIS 3.2.2 (Team, 2021), e a linguagem R (Core Team R Development, 2017).

2.2.2.2 Potenciais preditores do modelo

Com o objetivo de representar a produtividade de soja a nível local, geramos um modelo matemático preditivo de produtividade agrícola com os dados das 63 propriedades rurais para calibração, e com variáveis de sensoriamento remoto (Tabela 1; Anexo 7) escolhidas em função da sua importância biofísica para o desenvolvimento da soja. Foram consideradas o Índice de Vegetação Melhorado (EVI) conhecido como um *proxy* para a produtividade agrícola, e cinco variáveis climáticas para entender a variabilidade da soja em função das condições climáticas locais, considerando que a vegetação é capaz de influenciar o clima em diferentes escalas. Assim, os potenciais preditores testados foram o percentil de 90% do EVI, déficit hídrico climático máximo (CWD), período de seca máximo (DS), número total de dias secos (NDD), déficit de pressão de vapor máximo (VPD) e temperatura diurna máxima da superfície (DLST).

Tabela 1. Descrição dos potenciais preditores para o modelo de produtividade agrícola. Para cada variável foi considerado um mês da estação de crescimento da soja (outubro a março).

Potenciais preditores (mensal)	Escala espacial/temporal	Origem	Produto*
$EVI_{percentil\ 90\%}$	500 m/diário	MODIS Terra Daily	MOD09GA_006_EVI
$DLST_{máxima}$	1 km/diário		MOD11A1.061
$VPD_{máximo}$	4,5 km/mensal	TerraClimate	IDAHO_EPSCOR/TERRACLIMATE
$CWD_{máximo}$			
NDD_{total}	5,5 km/diário	CHIRPS	UCSB-CHG/CHIRPS/DAILY
$DS_{máximo}$			

*O nome de cada produto é referente ao que está no repositório do *Google Earth Engine*.

Todas as variáveis foram extraídas pela média dos polígonos de soja apresentados na Figura 1 (polígonos em roxo) para cada uma das propriedades rurais, e com resolução temporal determinada pelo tempo mínimo de início de plantio ao início de colheita (Anexo 5) de cada propriedade quando disponível. O tempo mínimo foi escolhido a fim de excluir áreas que iniciaram o plantio tardiamente e áreas que estariam em momento de dessecação ou colheita da soja. Dessa forma, considerando o estágio mínimo da estação da soja para a nossa área de estudo, cada variável foi obtida mensalmente entre outubro e março, em que se espera a fase de semeadura e emergência da soja entre outubro e dezembro, e as fases de desenvolvimento vegetativo, formação de grãos e início de maturação entre novembro e março (Evangelista et al., 2016; Farias et al., 2007; Fehr and Caviness, 1977). Todos os dados de sensoriamento remoto dessa etapa foram obtidos pelo *Google Earth Engine* (GEE, 2021; Gorelick et al., 2017).

2.2.3 Análise estatística

2.2.3.1 Escolha dos preditores

Para identificar os fatores compostos de rendimento da cultura de soja realizamos uma seleção automatizada de modelos através do método *stepwise* com auxílio de regressão linear múltipla (Alves et al., 2014; Ribeiro et al., 2019) pela a função *regsubsets* do pacote *leaps*. O método *stepwise* testa todas as possibilidades de combinações entre variáveis e retorna a combinação que melhor explica a variabilidade da variável dependente. Para isso, os dados foram previamente padronizados por meio do cálculo de *Z-score* (equação 1), em que x é o valor observado, μ é a média amostral e σ é o desvio padrão amostral.

$$Z = \frac{x-\mu}{\sigma} \quad \text{(eq. 1)}$$

O modelo global incluiu seis diferentes variáveis preditoras listadas na Tabela 1 (EVI, DLST, VPD, CWD, NDD e DS). Todas as variáveis foram derivadas de sensoriamento remoto com escala de tempo mensal na janela de cultivo da soja (outubro a março), totalizando 36 potenciais unidades preditoras (6 preditores por 6 meses) para todas as propriedades rurais em conjunto. Consideramos para o nosso modelo preliminar o ajuste por regressão linear múltipla (LM; equação 2) que estabelece uma relação entre a variável resposta (ŷ) e uma ou mais variável preditora (X_n), em que β diz o quão influente é a variável preditora, e o a descreve o intercepto, posição da linha quando o valor de X é zero:

$$\hat{y} = a + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad \text{(eq. 2)}$$

Consideramos como melhor modelo aquele com a combinação de variáveis que apresentou o maior R^2 ajustado ($R^2_{aj.}$). O $R^2_{aj.}$ (equação 3) penaliza o modelo com a adição de novas variáveis, e pode ser calculado por:

$$R^2_{aj.} = 1 - \frac{n-1}{n-(k+1)}(1 - R^2) \text{ (eq. 3)}$$

Em que *n* é o número de observações, *k* é o número de variáveis independentes envolvidas no modelo, e R^2 é o coeficiente de determinação (equação 4) obtido por:

$$R^2 = 1 - \frac{SQR}{STQ} \text{ (eq. 4)}$$

Sendo SQR a soma dos quadrados dos resíduos, e STQ a soma total dos quadrados.

O ajuste do melhor modelo para obter o modelo final foi feito com um ajuste de modelos lineares de efeitos mistos (LMM; equação 5), considerando as 63 propriedades rurais como fator aleatório, resultando em um modelo de rendimento agrícola adequado para toda a região de extrapolação da Figura 1 (área de linha pontilhada). Com o LMM obtemos os coeficientes de determinação marginal (R^2m) e condicional (R^2c) com que podemos interpretar a proporção da variação explicada pelo modelo. Nessa etapa utilizamos a função *lmer* do pacote *lme4*.

$$\hat{y} = a + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + propriedade \text{ (eq. 5)}$$

2.2.3.2 Validação do modelo e extrapolação

Para avaliar o desempenho do modelo preliminar (LM), realizamos validações cruzadas. Neste método, uma observação é removida da amostra total para o ajuste do modelo, e, em seguida, esta observação removida é utilizada para validação (Ribeiro et al., 2019). Essa análise foi feita de duas maneiras: validação cruzada "*leave one sample out*", em que todas as amostras foram retiradas para treinamento, e validação cruzada "*leave one farm out*", como validação espacial do modelo, em que todos os dados de cada fazenda foram retirados uma vez para o treinamento.

A qualidade do ajuste do esquema de validação cruzada "*leave one sample out*" foi testada pela análise da variância explicada em termos de $R^2_{aj.}$ (Wilks, 2006; Ribeiro et al., 2019). A validade do segundo esquema de validação cruzada foi avaliada através da média dos $R^2_{aj.}$ de cada observação (fazenda) que não foi considerada e que foi utilizada para validação cruzada. Além disso, foram incluídas outras medidas de acurácia, como o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE) (Tabela 2).

Tabela 2. Equações do viés, raiz do erro quadrado médio (RMSE) e erro absoluto médio (MAE), com o $\hat{y}$ representando o valor da produtividade de soja predita, y o valor das observações e N o número amostral.

Erro	Equação
Viés	$\frac{1}{N}\sum(\hat{y} - y)$
RMSE	$\sqrt{\frac{1}{N}\sum(\hat{y} - y)^2}$
MAE	$\frac{1}{N}\sum\lvert\hat{y} - y\rvert$

Adicionalmente, para validar nosso modelo de produtividade, também realizamos comparações entre os valores de produtividade de soja previsto pelo nosso modelo com os dados de produtividade municipal do IBGE. Para isso, utilizamos um gráfico de *boxplot* em série temporal e um gráfico um-pra-um, e fizemos uma análise de *bias* ou viés (Tabela 2), que indica a superestimação ou subestimação do modelo pela diferença entre os valores preditos e observados.

A extrapolação espacial da produtividade da soja foi realizada com base no modelo final (LMM) para as áreas de cultivo de soja de sequeiro para dentro da região que abrange os estados de Mato Grosso, Goiás, Distrito Federal, e porção do MATOPIBA (Figura 1 e 4). Utilizamos os mesmos produtos de sensoriamento remoto obtidos pelo GEE, resultando num mapa de produtividade agrícola para todos os anos de dados disponíveis (anos de colheita de 2010 a 2017). Através de uma malha municipal, fizemos a média anual de produtividade para cada município presente dentro dos limites da Spera et al. (2016), e comparamos com os dados de produtividade de soja disponibilizados pelo IBGE.

2.3 Relação entre a produtividade agrícola e a vegetação nativa da paisagem

Para verificar a relação entre vegetação nativa presente na paisagem e a produtividade de soja, amostramos a produtividade agrícola de forma aleatória na região e avaliamos sua relação com a vegetação presente na paisagem do entorno das áreas cultivadas (Figuras 3 e 4). Mais especificamente, seguimos os seguintes passos (Anexo 2): a) construímos um mapa de vegetação com a classe de cobertura florestal utilizando o produto do MapBiomas (2022), e unimos com o mapa de produtividade agrícola predita gerado com base no modelo como detalhado acima; b) amostramos a produtividade de soja em 100 pontos aleatórios e repetimos este procedimento 20 vezes (total de 2000 pontos); c) para cada ponto construímos *buffers* com raios variando entre 10 e 50 km, de 5 em 5 km, para avaliar a escala na qual a proporção de vegetação estava mais correlacionada com a produtividade de soja (Figura 4); d) calculamos a média de produtividade de soja para o *buffer* de 10 km, e a proporção de vegetação nativa para todos os tamanhos de *buffers* e para cada ano no período entre 2010 a 2017; e) utilizamos uma regressão linear e regressão linear segmentada para verificar a existência da relação e avaliar os coeficientes de determinação. A seguir, detalhamos cada uma destas etapas.

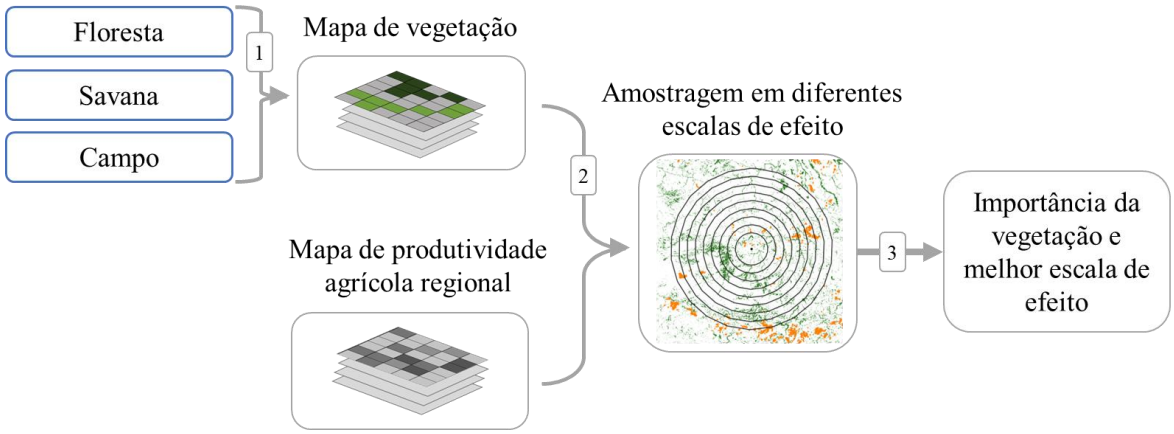


Figura 3. Esquema da segunda etapa do trabalho de verificação da importância da vegetação nativa para a produtividade agrícola de soja. 1 – Construção do mapa de vegetação pelos dados de cobertura do solo do MapBiomas (2022) para toda a área do mapa de produtividade agrícola predito; 2 – criação de 100 pontos aleatórios repetidos 20 vezes (2000 pontos), e de diferentes escalas de efeito em *buffers* de raios de 10km a 50km; 3 – amostragem da média de soja em 10km de *buffer* e da vegetação para todas as escalas e análises estatísticas.

2.3.1 Dados utilizados

A segunda fase da pesquisa está concentrada em partes da região do MATOPIBA e em todo o território dos estados de Mato Grosso, Goiás e Distrito Federal (Figura 1 e 4). A delimitação dessa área foi determinada pela disponibilidade de dados de cobertura do solo disponibilizados por Spera et al. (2016). Uma maior área de produtividade agrícola obtida a partir do nosso modelo aumentou a nossa capacidade de amostragem permitindo uma análise de médias de produtividade obtidas a partir do pixel. Além disso, foi possível estabelecer uma relação entre a vegetação e as variáveis de reflectância e clima que selecionamos como *proxy* para a produtividade da soja de sequeiro na nossa região.

2.3.2 Análise de sensoriamento remoto

2.3.2.1 Mapa de cobertura de vegetação e escala de efeito

O mapa de vegetação foi extraído para todo o perímetro da área de linha pontilhada da Figura 1, contemplando toda a área em que foi criado o mapa de produtividade agrícola a partir do nosso modelo. Para a elaboração deste mapa, utilizamos a classe floresta da Coleção 7 do Projeto Brasileiro Anual de Mapeamento de Uso e Cobertura da Terra (MapBiomas, 2022). O MapBiomas é disponibilizado na plataforma do *Google Earth Engine* (GEE) e apresenta séries temporais anuais de uso e cobertura da terra para todo o Brasil, com uma resolução de 30 m (*Landsat*) de 1985 a 2021 (Souza Júnior et al., 2020).

Avaliamos a relevância da floresta para a produção agrícola através de uma análise de regressão linear. Além disso, analisamos a escala de paisagem em que a vegetação nativa melhor se relaciona com a produtividade de soja. Para isso, realizamos uma amostragem regional a partir da junção do mapa de produtividade agrícola predita com o mapa de vegetação (Figura 4). Na área de soja, criamos 100 pontos aleatórios com 20 rodadas ou grupos de aleatorização, totalizando 2000 pontos aleatórios, e para cada ponto criamos *buffers* de círculos concêntricos de raios com distâncias de 10 a 50 km (variando de 5 em 5 km) em relação ao centro do ponto. O tamanho de *buffer* foi usado para identificar a escala na qual a vegetação nativa da paisagem estava mais relacionada à produtividade de soja predita.

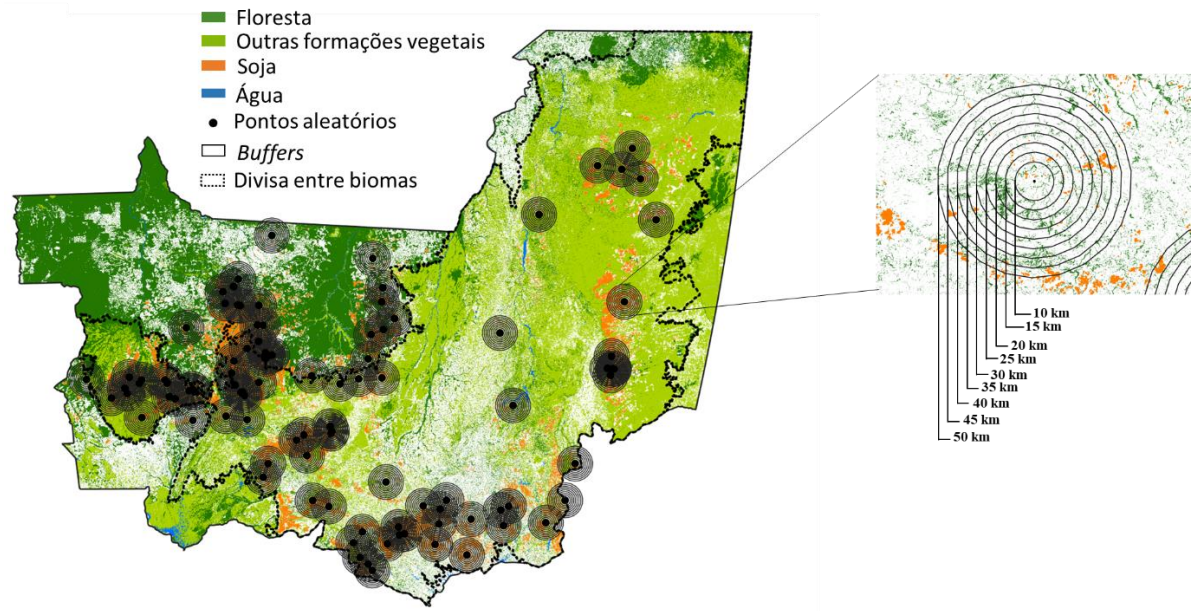


Figura 4. Ilustração de um dos grupos de randomização com 100 pontos aleatórios de amostragem sobre a área de extrapolação de produtividade de soja de sequeiro predita. Ao redor da área de soja estão as classes de floresta e outras formações vegetais para o ano de 2016 (MapBiomas, 2022). Para cada ponto foram gerados *buffers* com tamanho máximo de distância de 10 a 50 km (variando de 5 em 5 km) de raio em relação ao ponto central (esquema ao lado direito).

As amostragens relacionadas a produtividade da soja e cobertura do solo foram realizadas no GEE. Para a soja, nós amostramos a produtividade média predita dentro de um raio de 10 km de *buffer*. Para a vegetação, primeiro fizemos a reclassificação de cada classe de vegetação de interesse com valor igual a 1. Depois, utilizamos a função *ee.Reduce.sum* para somar a quantidade de pixels dentro da área de amostragem a fim de calcular a porção de área de cada *buffer* que é ocupada por vegetação.

A função de redução e soma do GEE utilizada é uma função de redução estatística sem peso que considera qualquer pixel com uma quantidade mínima de 0,5% de presença dentro da área de estudo. O pixel tem o seu valor calculado pela fração de sua presença, não sendo considerado como um pixel inteiro.

2.3.2.2 Precipitação

Para avaliar como a relação entre a produtividade da soja e a proporção de florestas na paisagem pode ser modulada pelo padrão de clima característico de cada região, amostramos a

precipitação para toda a série temporal para os anos de plantio (2009 a 2016) utilizando os dados do CHIRPS. Calculamos a precipitação média diária (mm) para todos os 2000 pontos aleatórios, e consideramos apenas a melhor escala de efeito da relação entre soja predita e proporção de floresta.

2.3.3 Análise estatística

Nas análises de importância da vegetação ao redor das áreas agrícolas para a produtividade da soja, nós consideramos a soja predita pelo nosso modelo e a área coberta por vegetação e a classe de floresta. Realizamos os testes para todo o perímetro da área de linha pontilhada da Figura 1 e para toda a série temporal dos anos de colheita entre 2010 e 2017.

Para testar nossa hipótese sobre a existência ou não de uma relação, nós realizamos uma análise de correlação para avaliar relações lineares considerando o coeficiente de correlação de Pearson (p). Consideramos como estatisticamente significativos resultados ao nível de 95%, com o valor de $p < 0,05$.

A melhor escala de efeito para a relação entre a produtividade da soja e a proporção de floresta na paisagem foi verificada por regressões lineares simples. Consideramos como a melhor escala aquela que apresentasse o maior R^2 e tendência de estabilização.

A partir da escala de efeito com maior R^2, modelamos regressões lineares segmentadas (equação 6) utilizando o pacote *segmented* do R, em que a variável independente é particionada em intervalos, e um segmento de linha separado é ajustado para cada intervalo, retornando o R^2. Considerando um único ponto de quebra na segmentação, o modelo segmentado pode ser expresso da seguinte forma:

$$\begin{aligned} \hat{y}_1 &= \beta_0 + \beta_1 x, \quad x \leq 0 \\ \hat{y}_2 &= \beta_0 + \beta_2 x + 0(\beta_0 - \beta_2), \quad x > 0 \end{aligned} \qquad \text{(eq. 6)}$$

Quando a função de segmentação não identificou um ponto de quebra, adotamos como ideal o modelo de regressão linear para a variável em questão. Para os modelos de vegetação nativa e de floresta, a precipitação foi considerada como uma variável de efeito aleatório.

3. Resultados

3.1 Modelo de produtividade agrícola

O modelo linear preliminar de produtividade agrícola com melhor ajuste apresentou poder de explicação de 54% ($R^2_{aj.}$ = 0,54), e incluiu os seguintes preditores (Figura 5; Anexo 8): índice de vegetação melhorado (EVI) para os meses de janeiro e março; temperatura máxima diurna da superfície da terra (DLST) para janeiro e fevereiro; déficit de pressão de vapor (VPD) para dezembro; número de dias secos (NDD) para novembro e outubro.

Os parâmetros usados para avaliar o ajuste dos modelos mostraram uma maior robustez do modelo para captar a variabilidade temporal dos dados do que a variabilidade espacial (Anexo 11). O coeficiente de determinação do modelo preliminar sem validação cruzada ($R^2_{aj.}$ = 0,54) foi ligeiramente maior do que o obtido por validação cruzada do tipo "*leave one sample out*" de 51% ($R^2_{aj.}$ = 0,51). No entanto, o coeficiente de determinação da validação cruzada do tipo "*leave one farm out*" foi cerca de 10% maior ($R^2_{aj.}$ = 0,64) em relação ao modelo final. O mesmo comportamento foi observado para os parâmetros de Erro Quadrático Médio (RMSE) e Erro Médio Absoluto (MAE) utilizados para avaliar a robustez dos modelos. Os maiores valores para estes parâmetros foram observados para o modelo de validação cruzada tipo "*leave one sample out*" (RMSE = 0,52 t/ha, MAE = 0,40 t/ha) em comparação ao modelo preliminar sem validação cruzada (RMSE = 0,50 t/ha, MAE = 0,39 t/ha). Os menores erros foram observados para validação cruzada do tipo "*leave one farm out*" (RMSE = 0,20 t/ha, MAE = 0,18 t/ha). Os resultados apresentados justificam a escolha de um modelo misto final (LMM), que considera as propriedades como fatores aleatórios no modelo. Apesar de o coeficiente de determinação do LMM (R^2 marginal) permanecer o mesmo do LM ($R^2_{aj.}$), houve uma significativa redução nos erros do modelo (Anexo 11).

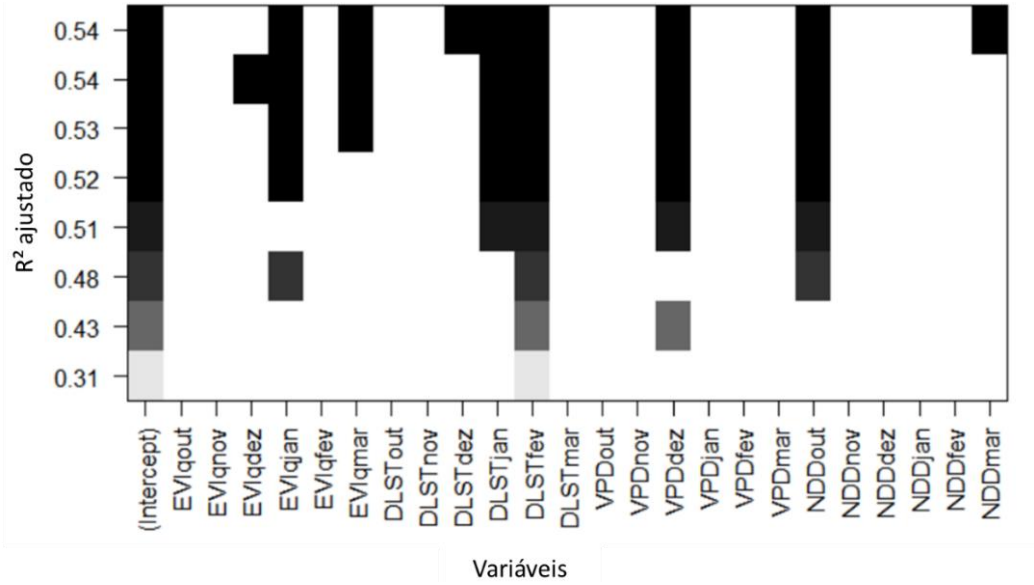


Figura 5. Gráfico de seleção de preditores pelo maior coeficiente de determinação ajustado (R^2 ajustado) obtido pelo método de seleção de *stepwise regression*.

O modelo final de produtividade agrícola (equação 7), que incluiu também as fazendas como fator aleatório (LMM), apresentou um poder de explicação de 71% (R^2condicional = 0,71), e os coeficientes reduziram para um RMSE igual a 0,38 t/ha e MAE igual a 0,28 t/ha (Anexo 11).

$$\hat{y} = 6.873795 + 0.803953EVI_{jan} + 1.152542EVI_{mar} - 0.023047DLST_{dez} - 0.041430DLST_{jan} - 0.055721DLST_{fev} - 0.420286VPD_{dez} - 0.010161NDD_{out} - 0.027909NDD_{mar} \quad \text{(eq. 7)}$$

O EVI foi positivamente relacionado com o aumento da produtividade agrícola, enquanto o DLST, VPD e NDD apresentaram uma relação negativa (equação 7; Figura 6; Anexo 9 e 10), sendo as variáveis mais importante do modelo o DLST de fevereiro e janeiro, o EVI de março e o VPD de dezembro. Os dados indicam que o aumento de 0,5 no valor do EVI para janeiro e março resulta em um aumento de 0,40 e 0,58 t/ha de soja, respectivamente. Aumentos de 5º C na temperatura diurna da superfície indicam diminuição de 0,12 t/ha quando em dezembro, 0,21 t/ha em janeiro e 0,28 t/ha em fevereiro; já aumentos de 0,50 no VPD no mês de dezembro resultaria em perdas de até 0,21 t/ha de soja, e o total de cinco dias sem chuva para outubro e março indica perdas de 0,05 t/ha e de 0,14 t/ha de soja.

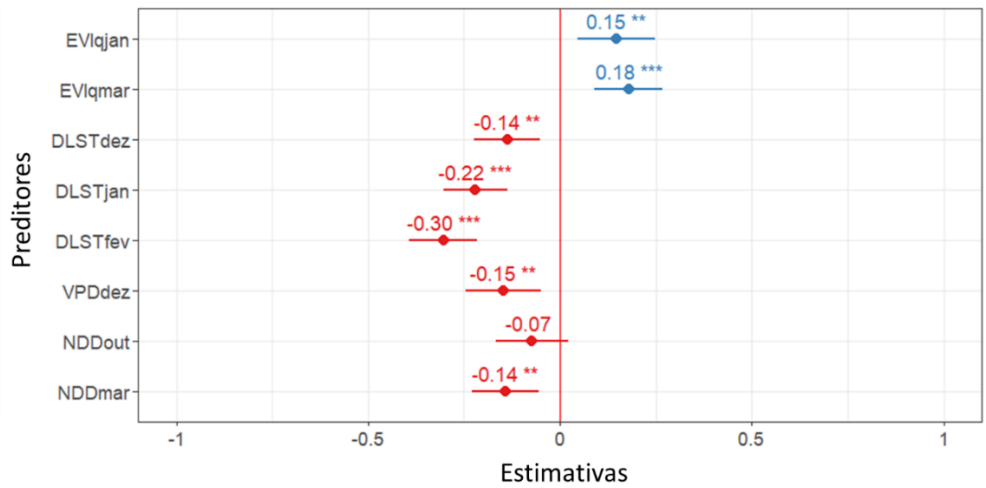


Figura 6. *Forest plot* dos coeficientes (valores sobre os pontos), erro padrão (comprimento das retas), e nível de significância (se **, $p \leq 0,05$; se ***, $p \leq 0,01$) dos preditores do modelo misto de produtividade agrícola (LMM). Valores previamente padronizados.

Nossa análise de validação utilizando os dados de produtividade municipal do IBGE indicou boa capacidade preditiva do modelo (Figura 7). O viés para a comparação com os dados do IBGE foi de - 0,13 t/ha considerando a série temporal completa, variando entre 0,04 t/ha para 2013 até - 0,34 t/ha em 2011 (Tabela 3).

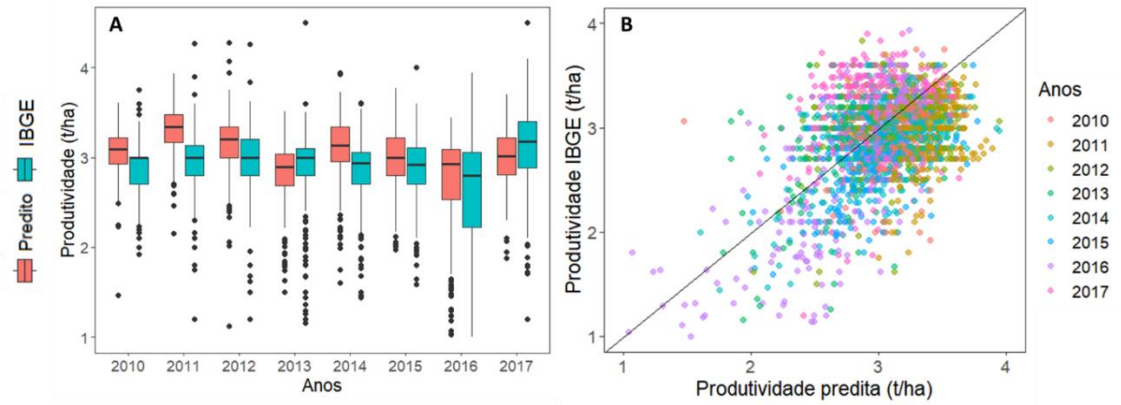


Figura 7. Comparação entre os dados de produtividade agrícola predito e observado (IBGE) a nível municipal utilizando *boxplot* (**A**) e gráfico um-pra-um da produtividade predita pela produtividade observada (IBGE), em que cada ponto é um município (**B**).

Tabela 3. Médias anuais e geral de produtividade de soja de sequeiro predita e observada (IBGE) calculadas a nível municipal para os anos de colheita entre 2010 e 2017, e valor de viés.

Ano	Produtividade de soja (t/ha)		Viés
	Predita	IBGE	
2010	3,07	2,89	-0,18
2011	3,33	2,99	-0,34
2012	3,17	3,01	-0,15
2013	2,84	2,88	0,04
2014	3,10	2,86	-0,25
2015	3,00	2,85	-0,16
2016	2,76	2,64	-0,12
2017	3,03	3,15	0,12
2010-2017	3,04	2,91	-0,13

3.2 Importância da vegetação nativa para a produtividade agrícola

A proporção de floresta presente na paisagem apresentou relação significativa ($p < 0,05$) com a produtividade de soja em todos os anos e aleatorizações, exceto em apenas três combinações para o ano de 2010 em *buffers* de 40 km, 45 km e 50 km, e para a média da série temporal em *buffers* de 10 km e 15 km (Anexo 12).

A formação florestal também apresentou efeito positivo e R^2 maior do que zero para todas as combinações (Anexo 13), com R^2 máximo de 50% para o ano de 2014. Um padrão semelhante foi observado para todas as escalas de efeito, sendo a escala de efeito de 30 km a que apresentou tendencia de estabilização, quando o R^2 variou entre 12% para o ano de 2010, e 34% para o ano de a 2014 (Figura 8).

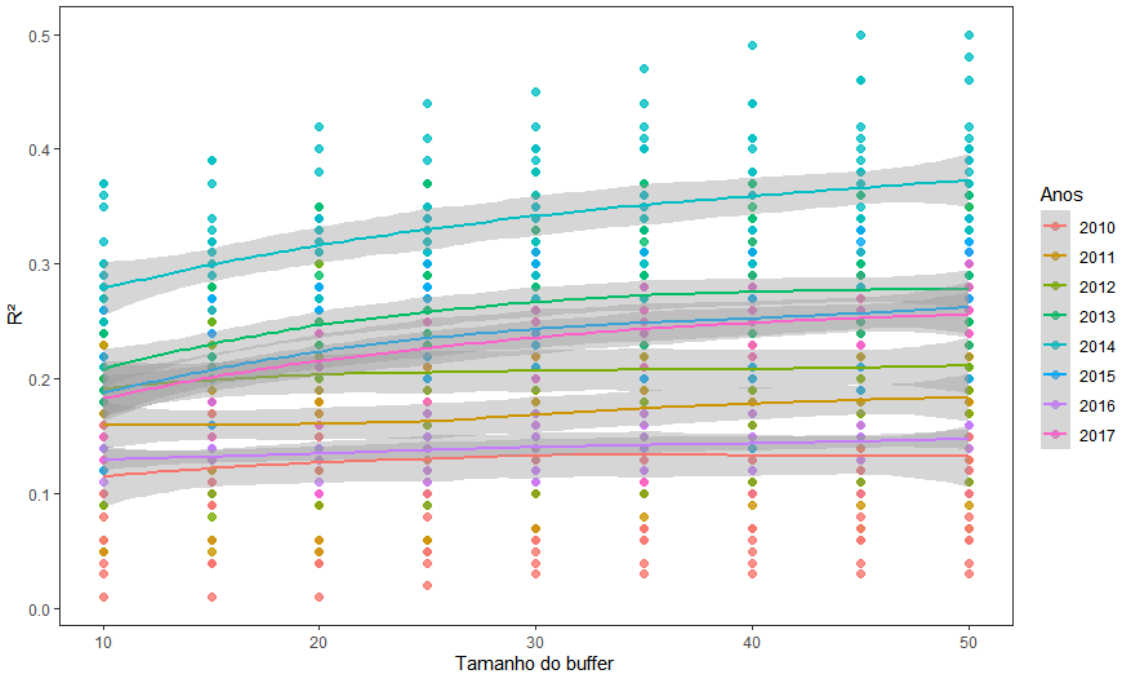


Figura 8. Coeficiente de determinação (R^2) da regressão entre soja predita e a proporção de floresta presente na paisagem. A soja predita foi amostrada pela sua média para *buffers* de 10 km de raio, e a vegetação pela sua proporção em *buffers* variando entre 10 e 50 km, sendo as amostras anuais para uma série temporal entre 2010 e 2017. Cada ponto representa a média de 20 grupos contendo 100 pontos gerados aleatoriamente sobre pixels de soja de sequeiro.

A cobertura de vegetação nativa na paisagem, considerando todas as formações vegetais na escala de efeito de 30 km, apresentou bom ajuste pela regressão linear não segmentada. A cobertura de vegetação nativa foi positivamente relacionada à produtividade de soja, com poder de explicação de 5% (R^2m = 0,05; Figura 9A; Tabela 4), indicando que aumentos na proporção de vegetação nativa da paisagem resulta em maior produtividade de soja, o que se repete para a maioria dos anos avaliados, exceto para safra 2015/16 (Anexo 14).

Quando isolamos apenas a classe de floresta para a mesma escala amostral, considerando toda a série temporal (2010 – 2017) o poder de explicação passa a ser de 30% (R^2m = 0,30; Figura 9B). Para essa variável o ajuste por regressão segmentada se mostrou adequado, aumentando a explicação da produtividade predita pela floresta para 40% (R^2m = 0,40), apresentando um ponto de quebra em 18% de cobertura florestal na paisagem. Tanto a regressão linear quanto a regressão segmentada apresentaram relações positivas. No caso da regressão segmentada, tanto o primeiro quanto o segundo *slope* foi positivo, com maior *slope* para o primeiro segmento (Figura 9B).

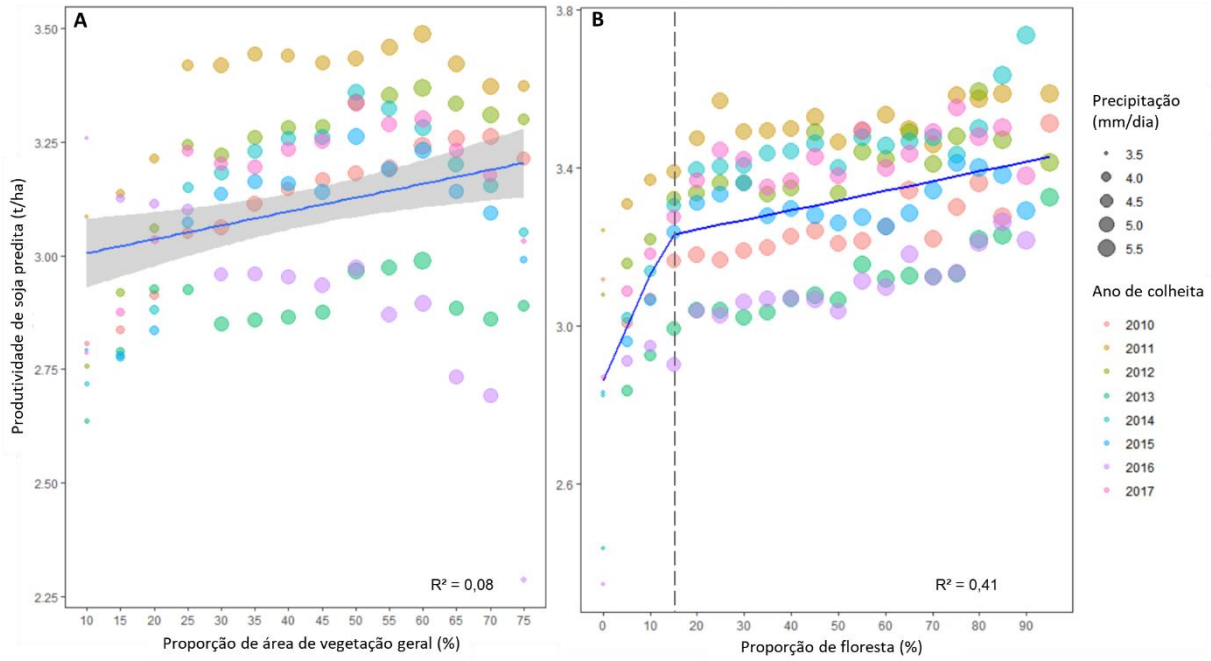


Figura 9. Produtividade de soja predita (t/ha) explicada (R^2) pela proporção de área coberta por vegetação (%) (A) e pela proporção de área coberta por floresta (%) (B) presente na paisagem agrícola em um raio de 30 km de *buffer*.

Tabela 4. Sumário dos modelos lineares mistos (LMM) utilizados para avaliar a precipitação diária média como variável aleatória nos modelos que verificam o efeito da vegetação nativa e da floresta sobre a produtividade de soja predita.

	Coeficientes	**Erro padrão**	**T**
Vegetação nativa LMM$_1$: R^2m = 0,05; R^2c = 0,47			
Intercepto	3,006700	0,049294	60,995
Vegetação nativa (%)	0,002472	0,001031	2,397
Floresta LMM$_2$: R^2m = 0,30; R^2c = 0,44			
Intercepto	3,0830515	0,0294226	104,785
Floresta (%)	0,0043135	0,0005884	7,331
Floresta LMM segmentada: R^2m = 0,40; R^2c = 0,44			
Intercepto	2,8597323	0,056299	50,79515
Floresta (%)	0,0275101	0,008590	3,20255

Estimativa = intercepto ou *slope* da variável preditora no modelo; R^2m = coeficiente de determinação marginal; R^2c = coeficiente de determinação condicional.

A regressão segmentada indicou fortes reduções na produtividade de soja quando a proporção de florestas na paisagem reduz para níveis menores que 26%. Os maiores *slopes*

foram observados para o primeiro segmento para todos os anos estudados (Anexo 15 e 16), com ponto de quebra máximo da proporção de floresta de 26% em 2011 e mínimo de 5% em 2010, indicando forte redução na produtividade em paisagens com menores coberturas de vegetação. O *slope* da regressão no segundo segmento foi positivo para a maioria dos anos avaliados, indicando que o aumento na quantidade de vegetação nativa da paisagem resulta em maior produtividade de soja. O R^2 variou entre 0,16 e 0,54. Os anos que apresentaram maior importância foram os de 2013 (R^2 = 0,54), 2014 (R^2 = 0,52) e 2015 (R^2 = 0,40), com pontos de quebra de 9,28%, 18,90% e 21,17%.

A produtividade média anual apresentou forte correlação com os *slopes* 1 e 2 e o ponto de quebra da regressão segmentada (Figura 11). Os *slopes* 1 e 2 (Figura 11A e B) apresentaram correlação negativa com a produtividade, indicando maior contribuição da floresta para a produtividade em anos com condições ambientais menos adequadas. Por outro lado, o ponto de quebra (Figura 11C) apresentou relação positiva com a produtividade média de soja.

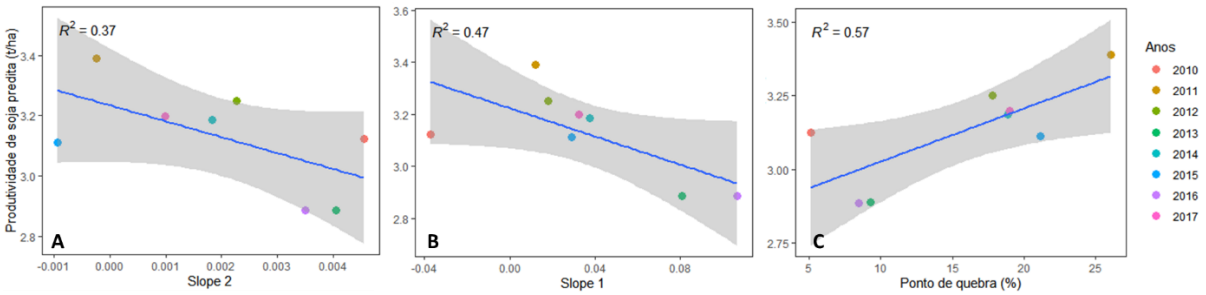


Figura 10. Correlações para a média anual de produtividade de soja predita com as métricas obtidas pela regressão linear segmentada. *Slope* 1, reta que antecede o ponto de quebra (A). *Slope* 2, reta que procede o ponto de quebra (B). Ponto de quebra, porcentagem de importância máxima de proporção de floresta dentro de um buffer de 30 km de raio para a produtividade agrícola (C).

4 Discussão

4.1 Modelo de produtividade agrícola: escolha dos preditores e performance do modelo

Nosso modelo mostrou boa capacidade de predição da produtividade de soja, e pode ser usado para compreender os efeitos de variações climáticas sobre a sua produtividade. O modelo final explica 71% da variância da produtividade de soja da região (R^2 = 0,71) e viés bem distribuídos no espectro de produtividade agrícola. O modelo pode ser considerado robusto, pois combina o índice de reflectância dos campos cultivados com fatores climáticos que estão

relacionados às primeiras fenofases de desenvolvimento da soja, o que torna o modelo potencialmente útil para a previsão da produtividade antes da colheita. Além disso, foi possível observar boa capacidade de captura da variabilidade espacial e temporal na produtividade da soja para a região estudada. Dessa forma, o uso do modelo pode gerar informações preditivas relevantes para decisões de mercado e antecipação dos impactos de condições ambientais e climáticas que levam a reduções na produtividade da soja.

Em nossos testes, o EVI foi crucial para a construção do modelo de produtividade de soja. Através do EVI é possível acompanhar os picos de produtividade agrícola ao longo da série temporal, e a sua maior relevância no ano de seca demonstra a capacidade de capturar a resposta espectral da cultura às mudanças nas condições de clima. O EVI, em geral, está relacionado às variações das condições hídricas da área de cultivo, e tende a diminuir em períodos de menor precipitação (Schaparini et al., 2020). As informações obtidas através do EVI refletem a dependência da disponibilidade de água para a produção de fotoassimilados, uma vez que o metabolismo da planta é afetado em situações de estresse hídrico durante toda ou em parte da estação de crescimento, o que reduz os rendimentos da colheita.

A influência das variáveis climáticas no modelo foi consistentemente observada, corroborando as expectativas iniciais. O Déficit de Pressão de Vapor (VPD), a Temperatura da Superfície da Terra (LST) e o Número de Dias Secos (NDD) apresentaram um impacto negativo significativo na produtividade agrícola da soja à medida que suas intensidades aumentaram. Essas variáveis climáticas afetam as condições ideais de temperatura e precipitação (Silvério et al., 2015) do agroecossistema, que são essenciais para o crescimento da cultura, resultando em maior estresse das plantas e, consequentemente, em menor produção de grãos (Bowles et al., 2018). Por exemplo, as altas temperaturas podem influenciar a disponibilidade de nutrientes para as plantas e comprometer a qualidade das sementes. Além disso, o aumento do VPD leva a uma maior demanda atmosférica por água, o que é agravado em combinação com um aumento no número de dias secos (Rattis et al., 2021). Esses fatores climáticos, desempenham um papel crucial nas diferentes fases fenológicas da cultura da soja, podendo determinar o sucesso ou fracasso da produção.

4.2 Importância da vegetação para a produtividade agrícola

A análise da escala de efeito mostrou que as florestas no entorno das lavouras de soja têm maior correlação quando consideramos *buffers* de 30 km, o que pode estar ligado às

variações na temperatura entre as florestas e áreas desmatadas, além do calor sensível e da circulação do ar. A temperatura das florestas é influenciada pelas variações na temperatura de áreas desmatadas até uma distância de até 50 km (Cohn et al., 2019), a escala de efeito semelhante ao encontrado no nosso estudo. Cohn et al. (2019) atribuíram a escala à capacidade de transporte da temperatura do ar e ao aquecimento do ar por processos de convecção nas áreas desmatadas.

Encontramos em nosso estudo uma relação positiva significativa entre a produtividade de soja e a proporção de floresta na paisagem circundante da área plantada. Isso evidencia as importantes contribuições dos ecossistemas naturais para a produção agrícola e destaca a relevância da preservação dessas áreas em benefício da sustentabilidade na produção de alimentos. A presença de florestas na paisagem proporciona uma série de serviços ecossistêmicos que podem influenciar positivamente a produtividade. Estes serviços incluem polinização, controle de pragas agrícolas por meio de inimigos naturais, e moderação das condições climáticas extremas nas lavouras (Shahzad et al., 2021). Particularmente, a forte redução na produtividade de soja em paisagens com menos de 30% de vegetação nativa, como o demonstrado em nosso estudo, evidencia possível limiar associado aos serviços ecossistêmicos, como a polinização, uma vez que a cultura da soja é fortemente dependente da polinização de insetos como as abelhas (Gazzoni, 2017). Dessa forma, a presença de florestas na paisagem aumenta a diversidade e a abundância desses polinizadores, melhorando a polinização da cultura e, consequentemente, a produtividade.

As florestas também desempenham um papel importante na moderação do clima local. As florestas contribuem para a regulação da temperatura, reduzindo os extremos de calor nas áreas cultivadas. Essas condições climáticas mais favoráveis podem favorecer o desenvolvimento saudável da cultura de soja e aumentar sua produtividade (Leite-Filho et al., 2021). Áreas desmatadas podem apresentar temperaturas da superfície até 5°C mais altas do que as florestas e experimentar um fluxo de calor sensível maior, o que leva a um aumento da temperatura do ar (Silvério et al., 2015). Esse calor é transportado pelo vento e se dissipa por toda a paisagem local. Consequentemente, as florestas atuam como reguladoras de extremos de temperatura (Cohn et al., 2019), resultando em uma maior produtividade de soja em paisagens com uma ampla cobertura florestal.

Portanto, os resultados do nosso estudo enfatizam a importância da conservação e preservação dos ecossistemas naturais, como as florestas, no entorno das áreas agrícolas. A manutenção da proporção de floresta na paisagem pode contribuir para garantir a

sustentabilidade e a resiliência dos sistemas agrícolas, promovendo uma maior produtividade de culturas como a soja. Essas descobertas destacam a necessidade de políticas e práticas agrícolas que incentivem a conservação dos ecossistemas naturais em prol da produção agrícola sustentável.

A proporção de floresta nativa na paisagem parece ser ainda mais relevante para a produtividade de soja, o que demonstra que os serviços ecossistêmicos fornecidos pelas áreas de florestas podem contribuir para a redução de efeitos adversos de períodos de seca prolongada e baixa precipitação, além de prevenir grandes perdas na produção de soja. No nosso estudo, os *slopes* do primeiro e do segundo seguimento mostraram uma forte correlação com a produtividade média anual de soja, o que indica uma maior contribuição das florestas nativas para anos com condições climáticas menos adequadas. Isso mostra que a preservação dos remanescentes florestais e, consequentemente, os seus serviços ecológicos podem ajudar a conter os efeitos das instabilidades climáticas que estão cada vez mais frequentes devido ao aumento da mudança climática. Além disso, pode-se prevenir a migração das áreas agrícolas como já tem ocorrido nos últimos anos (Rattis et al., 2021).

5. Conclusão

O nosso modelo final, que incluiu variáveis climáticas e condições do campo de cultivo, permitiu prever a variação espacial e temporal na produtividade de soja para a região do MATOPIBA, Goiás e Mato Grosso. Em estudos futuros sugerimos a ampliação para novas áreas a fim de testar a robustez do modelo e de gerar calibrações necessárias para a adequação em diferentes regiões.

A vegetação influencia diretamente a produtividade de soja das áreas cultivadas, com grandes reduções na produtividade de soja quando a proporção de floresta em um *buffer* de 30 km está abaixo de 18%. Este estudo mostra que a preservação de remanescentes florestais como prestadores de serviços ecossistêmicos e climáticos é essencial para a produtividade agrícola na região do MATOPIBA, Goiás e Mato Grosso, que juntos somam como um dos maiores produtores agrícolas do mundo.

6. Referências bibliográficas

Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Alves, M.F., Lotufo, A.D.P., Lopes, M.L.M., 2014. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas 1, 1–6. https://doi.org/10.5540/03.2013.001.01.0144

ANA - Agência Nacional de Águas e Saneamento Básico, 2021. Atlas Irrigação. Uso da água na agricultura irrigada, Agência Nacional de Águas e Saneamento Básico.

Bieto, J.A., Talon, M., 1996. Fisiologia y bioquímica vegetal. Interamericana McGraw-Hill, Madrid.

Bowles, T.M., Atallah, S.S., Campbell, E.E., Gaudin, A.C.M., Wieder, W.R., Grandy, A.S., 2018. Addressing agricultural nitrogen losses in a changing climate. Nat. Sustain. 1, 399–408. https://doi.org/10.1038/s41893-018-0106-0

Chen, J., Saunders, S.C., Crow, T.R., Naiman, R.J., Brosofske, K.D., Mroz, G.D., Brookshire, B.L., Franklin, J.F., 1999. in Forest Microclimate and Ecosystem Ecology Landscape Variations in local climate can be used to monitor and compare the effects of different management regimes. Bioscience 49, 288–297.

Cohn, A.S., Bhattarai, N., Campolo, J., Crompton, O., Dralle, D., 2019. Forest loss in Brazil increases maximum temperatures within 50 km. Environ. Res. Lett. 14, 84047. https://doi.org/10.1088/1748-9326/ab31fb

Davidson, E.A., De Araüjo, A.C., Artaxo, P., Balch, J.K., Brown, I.F., Mercedes, M.M., Coe, M.T., Defries, R.S., Keller, M., Longo, M., Munger, J.W., Schroeder, W., Soares-Filho, B.S., Souza, C.M., Wofsy, S.C., 2012. The Amazon basin in transition. Nature 481, 321–328. https://doi.org/10.1038/nature10717

Evangelista, B.A., Silva, F.A.M. da, Simon, J., Campos, L.J.M., Vale, T.M. do, 2016. Zoneamento de risco climático para determinação de épocas de semeadura da cultura da soja na região MATOPIBA 1–23.

Farias, J.R.B., Nepomuceno, A.L., Neumaier, N., 2007. Ecofisiologia da soja. Circ. Técnica 48 - Embrapa Soja.

Fehr, W.R., Caviness, C.E., 1977. "Stages of soybean development" (1977). Special Report. 80. Iowa State Univ. 1–12.

Flach, R., Abrahão, G., Bryant, B., Scarabello, M., Soterroni, A.C., Ramos, F.M., Valin, H., Obersteiner, M., Cohn, A.S., 2021. Conserving the Cerrado and Amazon biomes of Brazil protects the soy economy from damaging warming. World Dev. 146. https://doi.org/10.1016/j.worlddev.2021.105582

Garibaldi, L.A., Steffan-dewenter, I., Winfree, R., Aizen, M.A., Bommarco, R., Cunningham, S.A., Kremen, C., Carvalheiro, L.G., 2013. Honey Bee Abundance. Science (80-.). 339, 1608–1611.

Gazzoni, D.L., 2017. Soja e abelhas. Embrapa, Brasília.

GEE, 2021. Google Earth Engine [WWW Document]. URL https://earthengine.google.com/ (accessed 10.17.21).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

IBGE, 2023. IBGE. Sidra: Banco de Tabelas Estatísticas. Disponível em: https://sidra.ibge.gov.br/home/ipca/brasil Acesso em: setembro de 2023.

Leite-Filho, A.T., de Sousa Pontes, V.Y., Costa, M.H., 2019. Effects of Deforestation on the Onset of the Rainy Season and the Duration of Dry Spells in Southern Amazonia. J. Geophys. Res. Atmos. 124, 5268–5281. https://doi.org/10.1029/2018JD029537

Leite-Filho, A.T., Soares-Filho, B.S., Davis, J.L., Abrahão, G.M., Börner, J., 2021. Deforestation reduces rainfall and agricultural revenues in the Brazilian Amazon. Nat. Commun. 12. https://doi.org/10.1038/s41467-021-22840-7

MapBiomas, 2021. Projeto MapBiomas – Mapeamento Anual de Cobertura e Uso da Terra na Amazônia - Coleção 6 [WWW Document]. URL https://mapbiomas.org/ (accessed 10.17.21).

Miranda, E.E. de, Magalhães, L.A., Carvalho, C.A. de, 2014. Proposta de delimitação terrotiral do MATOPIBA. Embrapa 1–18.

Monasterolo, M., Musicante, M.L., Valladares, G.R., Salvo, A., 2015. Soybean crops may benefit from forest pollinators. Agric. Ecosyst. Environ. 202, 217–222. https://doi.org/10.1016/j.agee.2015.01.012

Mori, A.S., Lertzman, K.P., Gustafsson, L., 2017. Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. J. Appl. Ecol. 54, 12–27. https://doi.org/10.1111/1365-2664.12669

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Arai, E., Del Bon Espirito-

Santo, F., Freitas, R., Morisette, J., 2006. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. Proc. Natl. Acad. Sci. U. S. A. 103, 14637–14641. https://doi.org/10.1073/pnas.0606377103

Nogueira, E.M., Fearnside, P.M., Nelson, B.W., França, M.B., 2007. Wood density in forests of Brazil's "arc of deforestation": Implications for biomass and flux of carbon from land-use change in Amazonia. For. Ecol. Manage. 248, 119–135. https://doi.org/10.1016/j.foreco.2007.04.047

Nogueira, E.M., Nelson, B.W., Fearnside, P.M., França, M.B., Oliveira, Á.C.A. de, 2008. Tree height in Brazil's "arc of deforestation": Shorter trees in south and southwest Amazonia imply lower biomass. For. Ecol. Manage. 255, 2963–2972. https://doi.org/10.1016/j.foreco.2008.02.002

Oliveira, L.J.C., Costa, M.H., Soares-filho, B.S., 2013. Large-scale expansion of agriculture in Amazonia may be a no-win scenario. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/8/2/024021

Páscoa, P., Gouveia, C.M., Russo, A., Trigo, R.M., 2017. The role of drought on wheat yield interannual variability in the Iberian Peninsula from 1929 to 2012. Int. J. Biometeorol. 61, 439–451. https://doi.org/10.1007/s00484-016-1224-x

Rattis, L., Brando, P.M., Macedo, M.N., Spera, S.A., Castanho, A.D.A., Marques, E.Q., Costa, N.Q., Silverio, D. V., Coe, M.T., 2021. Climatic limit for agriculture in Brazil. Nat. Clim. Chang. https://doi.org/10.1038/s41558-021-01214-3

Ribeiro, A.F.S., Russo, A., Gouveia, C.M., Páscoa, P., 2019. Modelling drought-related yield losses in Iberia using remote sensing and multiscalar indices. Theor. Appl. Climatol. 136, 203–220. https://doi.org/10.1007/s00704-018-2478-5

Rodrigues, A.A., Macedo, M.N., Silvério, D. V., Maracahipes, L., Coe, M.T., Brando, P.M., Shimbo, J.Z., Rajão, R., Soares-Filho, B., Bustamante, M.M.C., 2022. Cerrado deforestation threatens regional climate and water availability for agriculture and ecosystems. Glob. Chang. Biol. 28, 6807–6822. https://doi.org/10.1111/gcb.16386

Schaparini, L.P., Fontana, D.C., Dalmago, G.A., Rodigheri, G., Fernandes, J.M.C., Santi, A., Vargas, P.R., 2020. Analise comparativa entre índices de vegetação e sua relação com o balanço hídrico em soja. Agrometeoros 27, 183–190. https://doi.org/10.31062/agrom.v27i1.26574

Shahzad, A., Ullah, S., Dar, A.A., Sardar, M.F., Mehmood, T., Tufail, M.A., Shakoor, A., Haris, M., 2021. Nexus on climate change: agriculture and possible solution to cope

future climate change stresses. Environ. Sci. Pollut. Res. 28, 14211–14232. https://doi.org/10.1007/s11356-021-12649-8

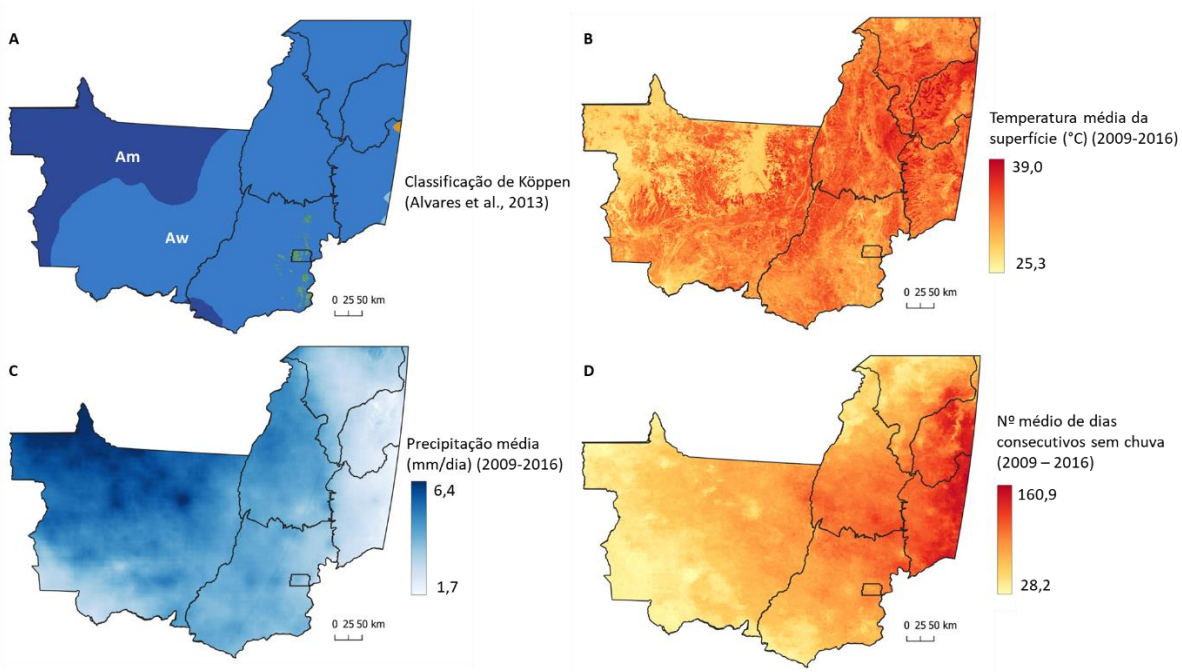
Silvério, D. V., Brando, P.M., Macedo, M.N., Beck, P.S.A., Bustamante, M., Coe, M.T., 2015. Agricultural expansion dominates climate changes in southeastern Amazonia: The overlooked non-GHG forcing. Environ. Res. Lett. 10. https://doi.org/10.1088/1748-9326/10/10/104015

Spera, S.A., Galford, G.L., Coe, M.T., Macedo, M.N., Mustard, J.F., 2016. Land-use change affects water recycling in Brazil's last agricultural frontier. Glob. Chang. Biol. 22, 3405–3413. https://doi.org/10.1111/gcb.13298

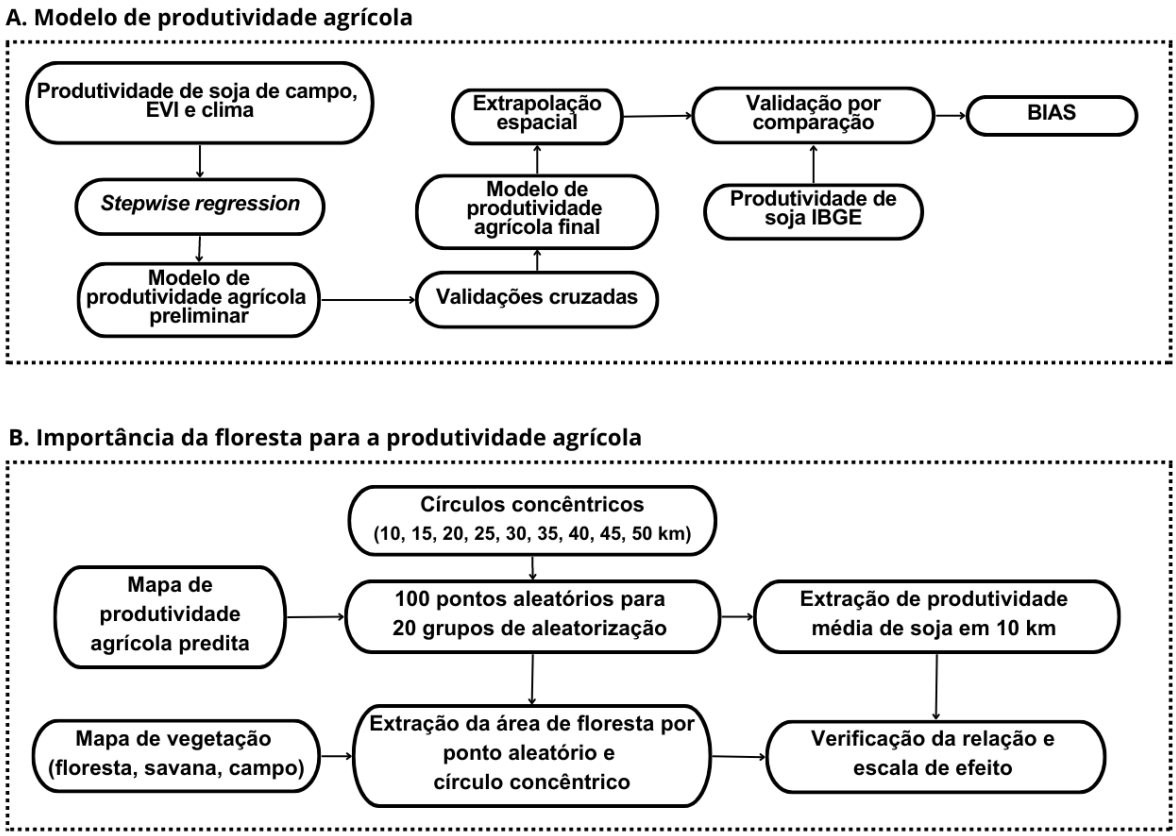
Yamada, T., 2005. The Cerrado of Brazil: A success story of production on acid soils. Soil Sci. Plant Nutr. 51, 617–620. https://doi.org/10.1111/j.1747-0765.2005.tb00076.x

7. Anexo

Anexo 1. Caracterização climática da região de estudo. Tipos climáticos (Am e Aw) predominantes de acordo com a classificação de Köppen (Alvares et al, 2013) (A), e médias para toda a série temporal de plantio (2009 – 2016) de temperatura diurna da superfície (ºC) (B), precipitação média diária (mm) (C) e número de dias consecutivos secos (precipitação abaixo de 3,3 mm) (D).



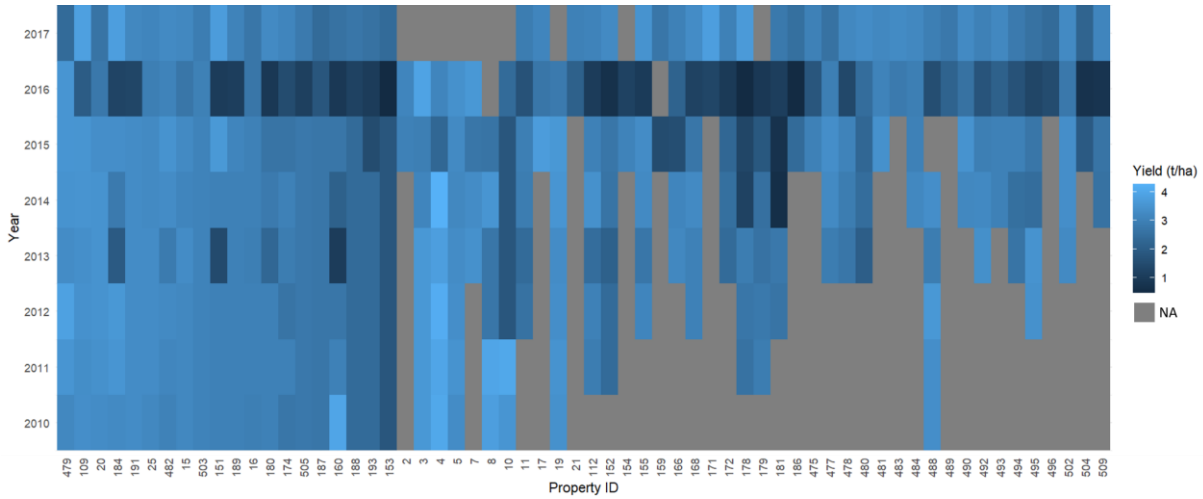
Anexo 2. Fluxo de trabalho detalhado para as duas etapas de (A) modelo de produtividade agrícola e (B) importância da floresta para a produtividade agrícola.



Anexo 3. Resumo dos dados de campo de produtividade agrícola de soja disponibilizados pelo Instituto de Pesquisa Ambiental da Amazônia (IPAM).

Estado	Município	Número de propriedades	Produtividade média (t/ha)
Bahia	Correntina	2	3,30
	Jaborandi	1	3,22
	Luís Eduardo Magalhães	4	3,25
	São Desidério	1	2,45
Maranhão	Alto Parnaíba	5	2,90
	Balsas	12	2,83
	Tasso Fragoso	5	2,76
Mato Grosso	Querência	1	3,44
Piauí	Baixa Grande	7	2,56
	Palmeiras	2	2,32
	Santa Filomena	6	2,51
	Sebastião Leal	3	2,61
	Uruçuí	7	2,30
Tocantins	Bom Jesus do Tocantins	1	3,14
	Campos Lindos	1	3,36
	Miracema	2	3,10
	Pedro Afonso	3	3,08

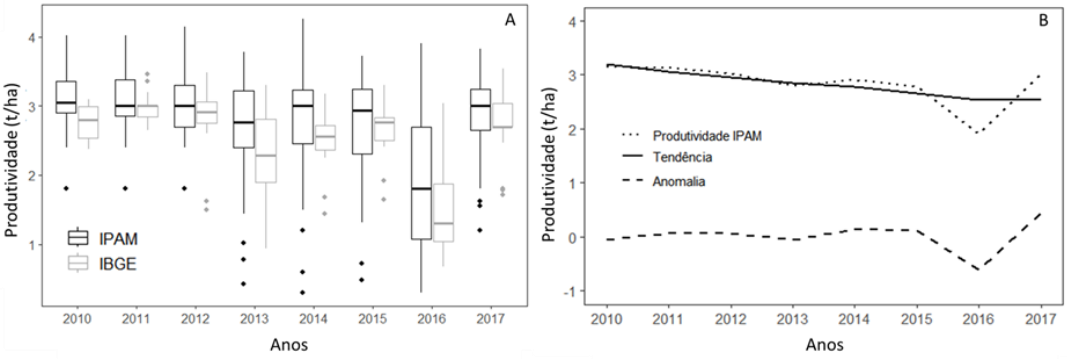
Anexo 4. Propriedades rurais utilizadas no modelo com e sem dados de produtividade agrícola de soja (retângulo em azul e retângulo em cinza) e produtividade média (t/ha) para cada ano safra.



Anexo 5. Comprimento máximo, mínimo e mediano da estação de crescimento da soja de 63 propriedades rurais do MATOPIBA e Mato Grosso.



Anexo 6. Dado de produtividade de soja de campo disponibilizado pelo Instituto de Pesquisa Ambiental da Amazônia (IPAM) e de seus respectivos municípios pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para os oito anos de colheita de 2010 a 2017 (A). Dado de produtividade de campo, linha de tendência e anomalia obtida pela retirada da tendência da soja (B).

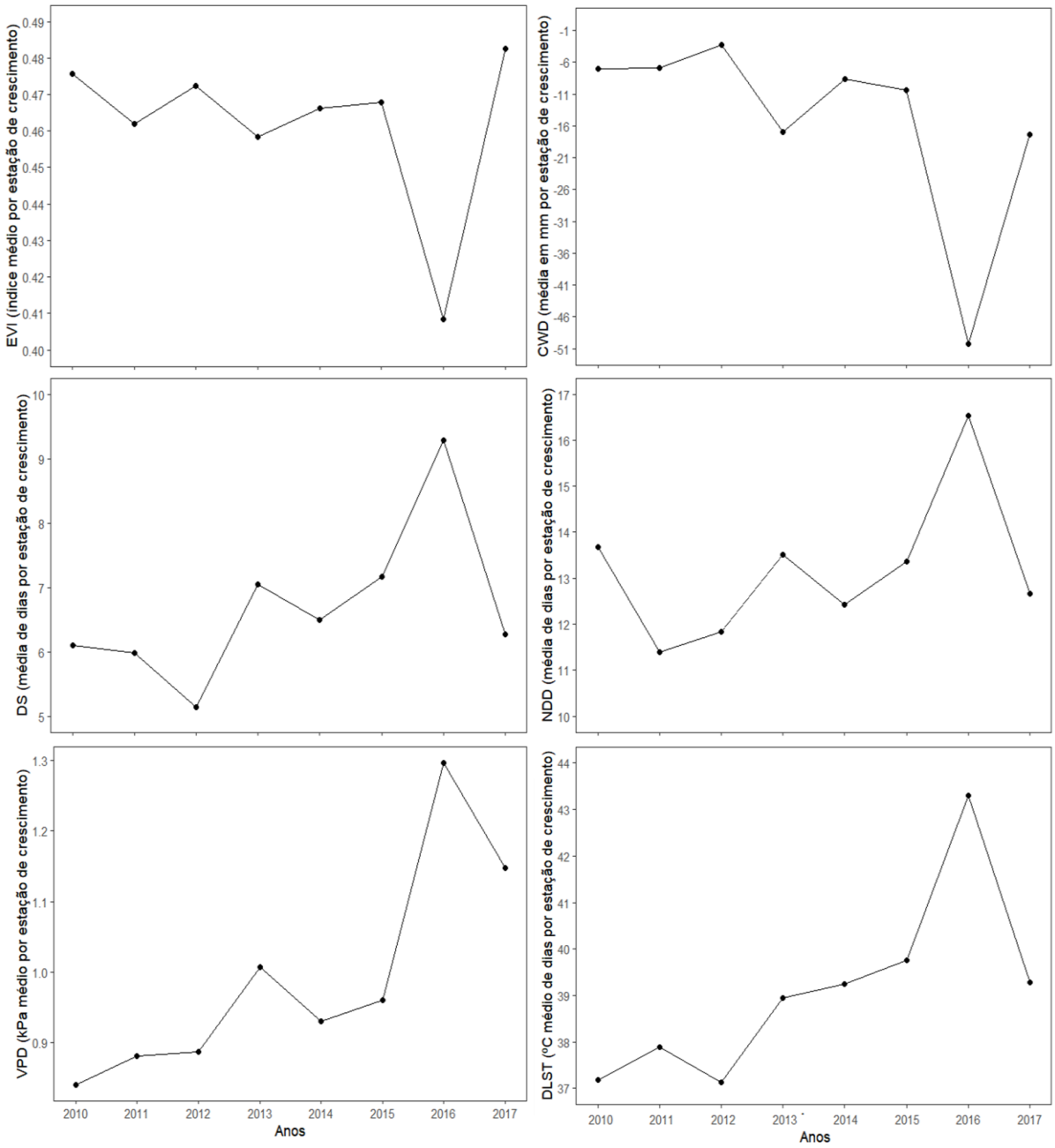


Os dados de produtividade disponibilizados pelo IPAM mostram comportamento semelhante com os dados do IBGE (Anexo 5A), com uma média total de produtividade de 2,55 t/ha e 2,68 t/ha, respectivamente, uma diferença de 0,13 t/ha entre as médias.

Em função de a produtividade observada das fazendas nessa série mostrar um declínio pela linha de tendência (Anexo 5B), nós decidimos continuar com os dados de produtividade observada ao invés da anomalia. Se a linha de tendência fosse positiva ao longo da série temporal, nós adotaríamos a hipótese de aumento da produtividade devido à intensificação agrícola e utilizaríamos a anomalia calculada pela retirada da tendência como uma variável dependente do modelo, mantendo assim apenas a influência do clima.

A produtividade de soja apresenta um coeficiente de variação médio anual de 22,5%, com valor máximo da série de 4,26 t/ha, e uma queda súbita de produtividade no ano de 2016 quando o menor valor de produtividade de toda a série (0,48 t/ha) foi observado. Essa queda na produtividade é acompanhada pelas variáveis escolhidas como potenciais preditores do modelo de produtividade agrícola (Anexo 6), bem como é esperado que aumentem ou diminuam a produtividade de soja a depender de seus comportamentos durante a estação de crescimento da soja.

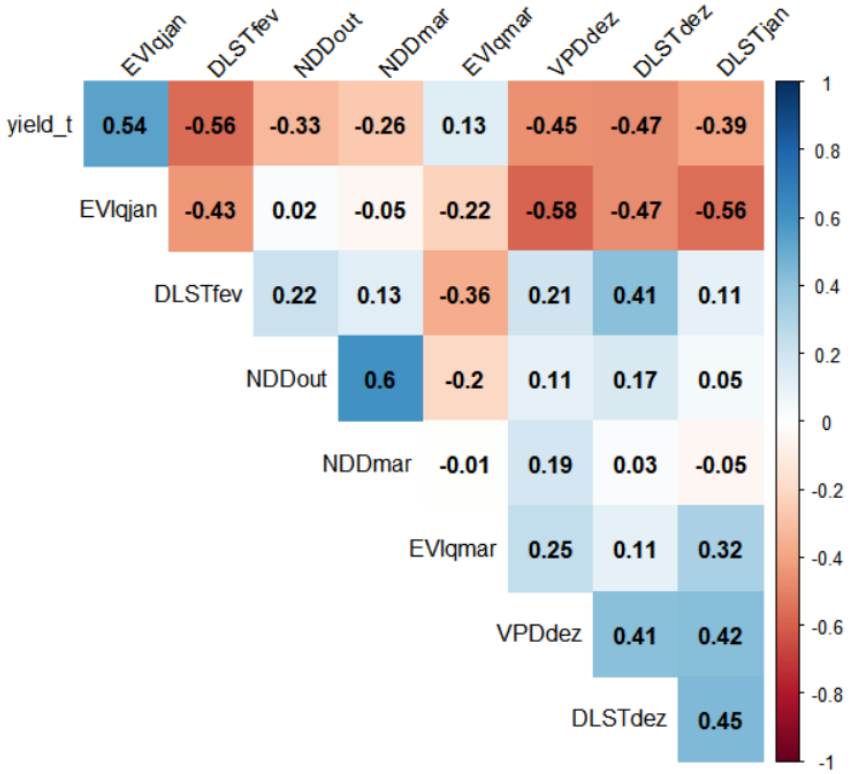
Anexo 7. Comportamento das variáveis escolhidas como potenciais preditores para o modelo de produtividade agrícola. As variáveis foram extraídas de dentro dos polígonos de soja selecionados para as propriedades rurais distribuídas no MATOPIBA e no Mato Grosso. As médias foram calculadas pela estação de crescimento da soja (outubro a março) considerando o ano agrícola. Os anos de 2010 a 2017 representam seus anos de colheita. Variáveis: quartil 90 do índice de vegetação aprimorada (EVI), *dry spell* ou período máximo de dias secos consecutivos (DS), número total de dias secos (NDD), déficit de pressão de vapor máximo (VPD), e temperatura diurna máxima da superfície da terra (DSLT).



Anexo 8. Resultados da regressão *stepwise* para a seleção dos preditores estatisticamente significativos ($p \leq 0,05$) para a produtividade de soja (*p* indica o número de preditores selecionados para cada modelo). Os preditores foram escolhidos pelos meses que indicam o maior coeficiente de determinação ajustado (R^2_{aj}). As variáveis de DS e CWD não foram selecionadas em nenhum dos melhores retornos da regressão.

Nome do modelo	*p*	EVI	DLST	VPD	NDD	$R^2_{aj.}$
Y1	7	jan, mar	jan, fev	Dez	out, mar	0,54
Y2	7	jan, dez, mar	jan, fev	Dez	out	0,54
Y3	6	jan, mar	jan, fev	Dez	out	0,53
Y4	5	jan	jan, fev	Dez	out	0,52
Y5	4	-	jan,fev	Dez	out	0,51
Y6	3	jan	Fev	-	-	0,48
Y7	2	-	Fev	Dez	-	0,43
Y8	1	-	Fev	-	-	0,32

Anexo 9. Correlação entre produtividade agrícola de soja de campo e preditores do modelo.



Anexo 10. Descrição do Modelo Linear Misto (LMM) de produtividade agrícola de soja. O modelo inclui dados de reflectância e clima para os meses da estação de crescimento adotada para a soja (outubro a março), e foi baseado em dados de produtividade de campo de parte da região do MATOPIBA e Nordeste Matogrossense.

	Coeficientes	**Coeficientes (*z-score*)**	**Erro padrão**	**T**
Intercepto	6.873795	0.00660	0.613347	11.207
EVIjan	0.803953	0.14668	0.282379	2.847
EVImar	1.152542	0.17892	0.287896	4.003
DLST dez	-0.023047	-0.13794	0.007212	-3.196
DLSTjan	-0.041430	-0.22007	0.007855	-5.274
DLSTfev	-0.055721	-0.30422	0.008250	-6.754
VPDdez	-0.420286	-0.14818	0.141913	-2.962
NDDout	-0.010161	-0.07314	0.006700	-1.517
NDDmar	-0.027909	-0.14135	0.008771	-3.182

EVI = Índice de Vegetação Melhorado; DLST = Temperatura Diurna da Superfície Terrestre; VPD = Déficit de Pressão de Vapor; NDD = Número de Dias Secos (< 3.3 mm).

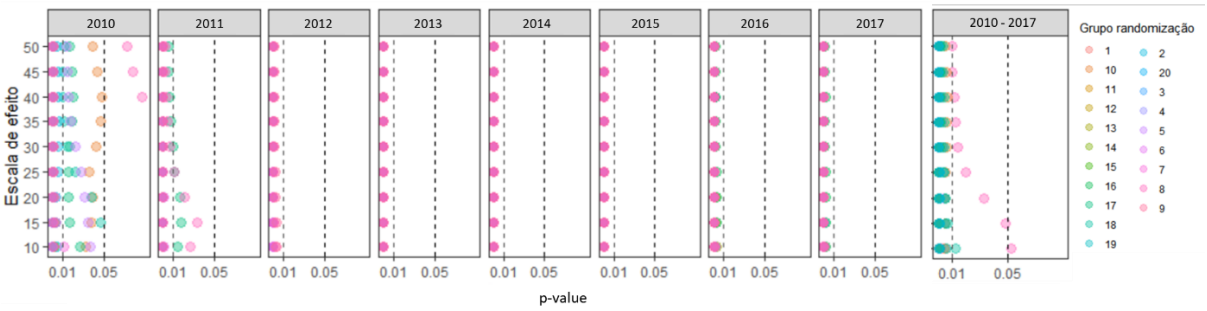
Anexo 11. Coeficientes para o modelo de produtividade agrícola preliminar, final e validações cruzadas.

Tipo de modelo	Origem	R^2	R^2 aj.	R^2c	RMSE	MAE
LM	Modelo preliminar	0,54	0,54	-	0,50	0,39
LM	Validação cruzada[1]	0,52	0,51	-	0,52	0,40
LM	Validação cruzada[2]	0,64	0,63		0,20	0,18
LMM	Modelo final	0,55	-	0,71	0,38	0,28

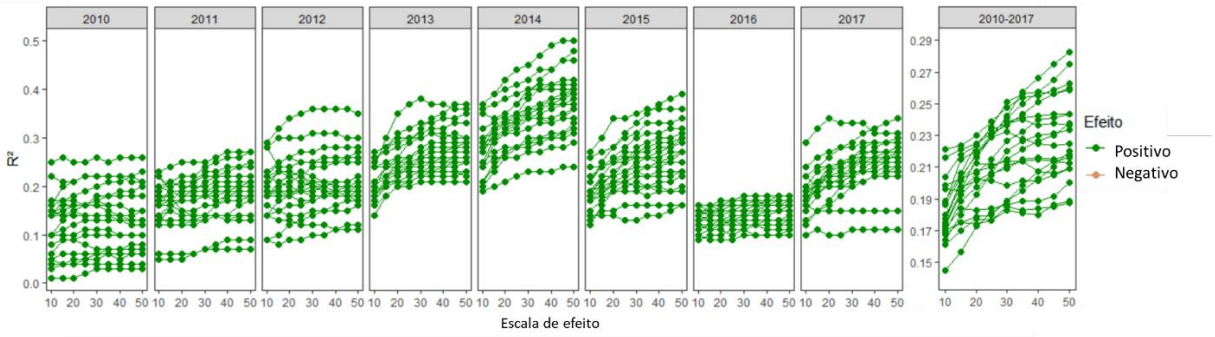
R^2 = coeficiente de determinação; R^2aj. = coeficiente de determinação ajustado; R^2c = coeficiente de determinação condicional; RMSE = Raiz do Erro Quadrado Médio; MAE = Erro Médio Absoluto.

[1] Validação cruzada *leave one sample out*; [2] Validação cruzada *leave one farm out.*

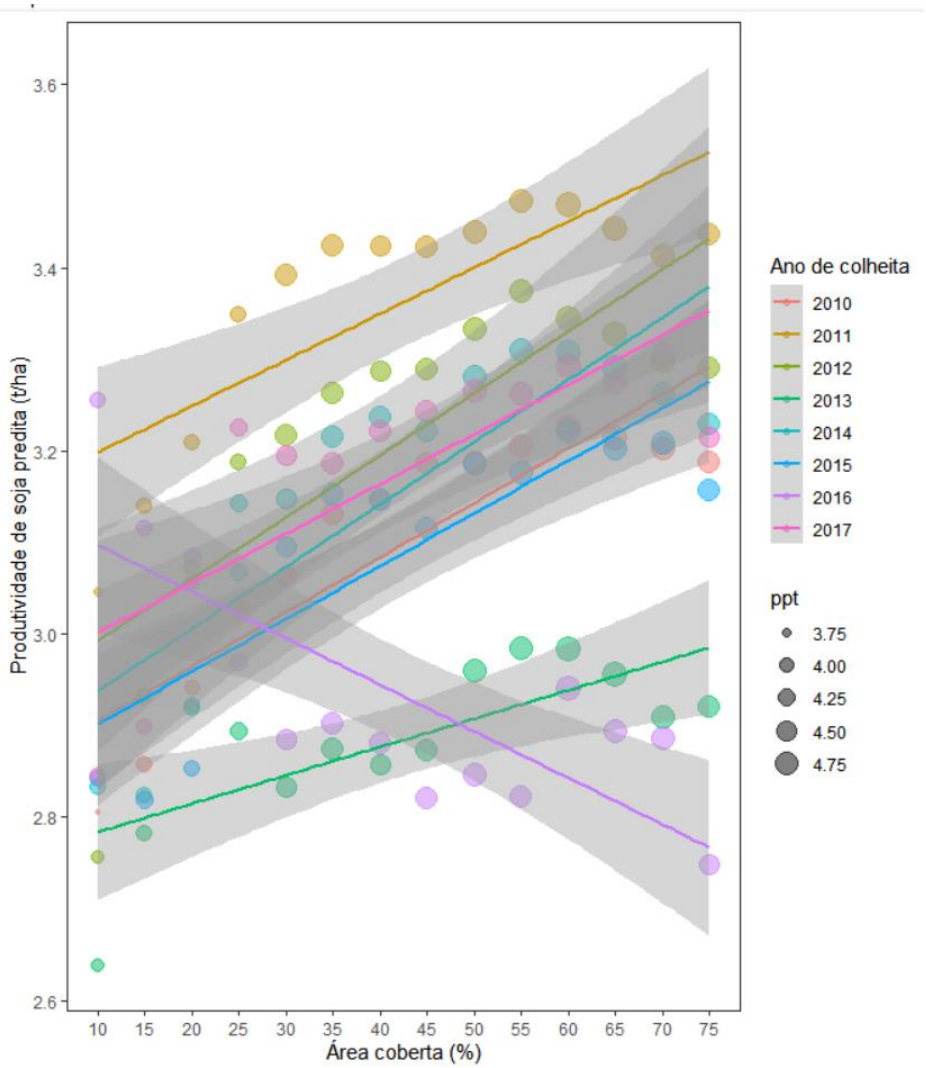
Anexo 12. p-value ($p < 0,01$ e $p < 0,05$) da relação entre produtividade de soja predita e floresta por escala de efeito (10 à 50km de buffer) para cada um dos 20 grupos de randomização dos 100 pontos aleatórios. Correlação obtida pela média de produtividade anual e da série temporal inteira dos anos de colheita (2010 – 2017).



Anexo 13. Efeito da relação (R^2) entre soja predita e floresta para Mato Grosso, Goiás, Distrito Federal, e partes da região do MATOPIBA. O efeito é dado pela interação da cobertura do solo com a produtividade de soja predita. O tamanho do *buffer* é referente aos círculos concêntricos com raios variando entre 10 e 50 km. Cada linha representa um grupo de aleatorização e a sua média para os 100 pontos aleatórios gerados para cada um.



Anexo 14. Relação entre produtividade de soja predita (t/ha) e área coberta por vegetação (%) por ano de colheita (2010 – 2017), com a quantidade média de precipitação diária (mm/dia) mostrada pelo tamanho dos pontos.



Anexo 15. Resumo dos modelos obtidos por regressão segmentada para a correlação entre produtividade de soja predita e floresta presente na paisagem agrícola.

Variável	Ano de colheita	Equações		Ponto de quebra (%)	R^2
		Se x ≤ z*	Se x > z		
Proporção de floresta	2010	ŷ =3,24469-0,05863x	ŷ=3,24469-0,05863z+(-0,05863+0,06308)(x-z)	3,245	0,16
	2011	ŷ=3.2158941+0,0117897x	ŷ=3,2158941+0,0117897z+(0,0117897-0,0121243)(x-z)	25,622	0,22
	2012	ŷ=3,031142+0,016378x	ŷ=3,031142+0,016378z+(0,016378-0,013901)(x-z)	17,713	0,25
	2013	ŷ=2,207651+0,096339x	ŷ=2,207651+0,096339z+(0,096339-0,091987)(x-z)	7,273	0,51
	2014	ŷ=2,753025+0,032509x	ŷ=2,753025+0,032509z+(0,032509-0,030694)(x-z)	19,634	0,47
	2015	ŷ=2,761762+0,026151x	ŷ=2,761762+0,026151z+(0,026151-0,026817)(x-z)	21,483	0,37
	2016	ŷ=2,134369+0,108281x	ŷ=2,134369+0,108281z+(0,108281-0,104617)(x-z)	7,371	0,31
	2017	ŷ=2,820906+0,029633x	ŷ=2,820906+0,029633z+(0,029633-0,028553)(x-z)	18,999	0,34
	2010-2017	ŷ=2,950910+0,018816x	ŷ=2,950910+0,018816z+(0,018816-0,016525)(x-z)	15,785	0,41

*x = porcentagem de vegetação dentro de 30 km de raio; z = ponto de quebra; ŷ = produtividade agrícola de soja

Anexo 16. Regressão linear segmentada entre soja predita (t/ha) e proporção de floresta (%) para a escala de efeito com maior R^2 (30 km). Cada ponto representa uma amostragem aleatória da média de soja predita dentro de uma escala de 10 km de buffer. A linha tracejada indica o ponto de quebra da regressão. A variação na coloração indica a precipitação diária média para toda a série temporal de plantio (2009 – 2016).

