

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FITOTECNIA

LACUNA DE PRODUTIVIDADE (“YIELD GAP”): ABORDAGEM MULTIVARIADA
PARA CLASSIFICAÇÃO DOS ATRIBUTOS QUE IMPACTAM A PRODUTIVIDADE
DA SOJA NO BIOMA CERRADO

Douglas Pedrini
Engenheiro Agrônomo/UFRGS

Dissertação apresentada como um dos requisitos
à obtenção do Grau de Mestre em Fitotecnia
Sistema de produção vegetal

Porto Alegre (RS), Brasil
Novembro de 2022

CIP - Catalogação na Publicação

PEDRINI, DOUGLAS
LACUNA DE PRODUTIVIDADE ("YIELD GAP"): ABORDAGEM
MULTIVARIADA PARA CLASSIFICAÇÃO DOS ATRIBUTOS QUE
IMPACTAM A PRODUTIVIDADE DA SOJA NO BIOMA CERRADO /
DOUGLAS PEDRINI. -- 2022.
80 f.
Orientador: ANDRÉ LUIS VIAN.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Faculdade de Agronomia, Programa de
Pós-Graduação em Fitotecnia, Porto Alegre, BR-RS,
2022.

1. Lacuna de produtividade. 2. Yield Gap. 3. Soja.
4. Cerrado. I. VIAN, ANDRÉ LUIS, orient. II. Título.

DOUGLAS PEDRINI
Engenheiro(a) Agrônomo(a) - UFRGS

DISSERTAÇÃO

Submetida como parte dos requisitos
para obtenção do Grau de

MESTRE EM FITOTECNIA

Programa de Pós-Graduação em Fitotecnia
Faculdade de Agronomia
Universidade Federal do Rio Grande do Sul
Porto Alegre (RS), Brasil

Aprovado em: 04/11/2022
Pela Banca Examinadora

ANDRÉ LUIS VIAN
Orientador(a)
UFRGS

SIMONE MUNDSTOCK JAHNKE
Coordenadora do Programa de
Pós-Graduação em Fitotecnia

CHRISTIAN BREDEMEIER
UFRGS

CARLOS ALBERTO BISSANI
Diretor da Faculdade de
Agronomia

CLAUDIR JOSÉ BASSO
UFSM - FW

TIAGO DE ANDRADE NEVES HORBE
CCGL

AGRADECIMENTOS

Agradeço à minha avó, Ramira, por ser a base do crescimento e sucesso de todos da família, e que, através de suas plantas na sacada de casa, é para mim um exemplo de amor à agricultura. Agradeço à minha mãe, Márcia, por sempre buscar o melhor para nós, sendo exemplo ímpar das recompensas obtidas da dedicação ao trabalho e da busca de conhecimentos. Ao Filipe, meu irmão e amigo, por ser exemplo de amor ao esporte, pelo companheirismo e pelas risadas, sou grato. Agradeço aos meus tios, Marcelo, Mauro e Mairo, que têm a curiosidade como virtude, e fazem disso propulsor para o constante aprendizado e me instigam a ser igual. Sou grato por estes que moldaram meu caráter.

Agradeço aos amigos Marquel e Cláudia, por quem tenho profunda admiração e respeito, e que tornaram possível minha trajetória profissional no sentido que tomou. Com eles aprendi e continuo aprendendo, sempre.

Ao Lucas e à Natali, grandes amigos que sempre estiveram dispostos a ajudar e por quem tenho grande felicidade de sempre poder contar, meu muito obrigado.

Ao Ronei, chefe, mentor, grande amigo e exemplo de ser humano, agradeço pelos ensinamentos, pela confiança e pelas oportunidades.

Agradeço ao grande amigo Renan Tavanti, cujo enorme conhecimento só é tão grande quanto sua humildade, sua bondade e seu bom coração. Sempre entendi o

conhecimento como algo inútil quando não compartilhado, e tenho Renan como exemplo de sucesso no uso de tudo o que sabe.

Ao professor Christian Bredemeier, cuja dedicação e empolgação em suas aulas sempre foi combustível para meu aprendizado, e cuja torcida pelo sucesso de seus alunos é inspiração para contribuirmos com entregas à agricultura brasileira, meu muito obrigado.

Ao meu querido orientador, André Vian, agradeço pelos ensinamentos, pelas conversas e principalmente pela compreensão no processo de desenvolvimento do trabalho, sei que terá sempre sucesso no seu caminho acadêmico, e que representa muito bem a nova geração de professores.

À UFRGS, instituição da qual tenho orgulho em pertencer, por ser grande, técnica, um templo e um lar, agradeço.

Por fim, agradeço a todos que acreditam na agricultura brasileira como fonte de alimentos, de empregos, de saúde e de riqueza para nosso país.

LACUNA DE PRODUTIVIDADE (“YIELD GAP”): ABORDAGEM MULTIVARIADA PARA CLASSIFICAÇÃO DOS ATRIBUTOS QUE IMPACTAM A PRODUTIVIDADE DA SOJA NO BIOMA CERRADO¹

Autor: Douglas Pedrini

Orientador: Prof. Dr. André Luis Vian

RESUMO

Diversos estudos buscam entender as lacunas de produtividade (Yield Gaps), identificando o quanto é possível aumentar a produtividade dos cultivos para atender a demanda crescente por alimentos. Estes estudos costumam utilizar modelagens complexas para estimativa das produtividades e dos Yield Gaps, e demonstram principalmente os valores de incremento possível, sem apresentar os principais atributos de manejo envolvidos no Yield Gap. O objetivo deste trabalho foi classificar o quanto a produtividade da soja é impactada pelos atributos de solo e de clima vinculados ao manejo da cultura, propondo processo de avaliação com base em estatística multivariada e de classificação de importância das variáveis. O estudo avaliou informações de 470 lavouras de produção de soja na safra 2021/2022 em 8 fazendas da SLC Agrícola nos estados do MA, BA, GO, MS e MT, e contou com 21 atributos vinculados a manejo da cultura. As produtividades reais variaram de 716,0 a 6044,0 kg ha⁻¹, e os atributos para cada talhão foram avaliados por comparativos de médias, demonstrando maiores produtividades nas fazendas Planalto e Pantanal no MS e Paladino na BA. As análises dos coeficientes de correlação e de componentes principais demonstraram forte correlação positiva da produtividade real com os valores de saturação de bases (V%) e pH do solo, bem como com teores de fósforo (P) e de boro (B), além da correlação negativa com a acidez potencial (H+Al) e a latitude (Y). A classificação não supervisionada retornou quatro grupos de talhões, distinguindo dois grupos de menores produtividades reais, com diferenças em suas produtividades potenciais e Yield Gap total, e outros dois grupos de altas produtividades, separados por atributos de fertilidade e teores de argila e nutrientes no solo. Utilizou-se Random Forest (RF) para modelar a produtividade da soja, obtendo-se ótimo ajuste para os valores de treino ($R^2 = 0,97$), proporcionando a avaliação da importância relativa das variáveis (IRV). Os valores de IRV demonstram que V%, H+Al, radiação solar, P e precipitação são os principais atributos que impactaram na produtividade para conjunto de dados. A quantidade e variabilidade dos dados permitiram que o fluxo de análise proposto retornasse bom resultado de identificação de atributos do Yield Gap de manejo.

¹ Dissertação de Mestrado em Fitotecnia, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (80f.) Novembro, 2022.

YIELD GAP: MULTIVARIATE APPROACH TO CLASSIFY IMPACT OF ATTRIBUTES IN SOYBEAN YIELD IN BRAZILIAN CERRADO BIOME¹

Author: Douglas Pedrini
Adviser: Prof. André Luis Vian

ABSTRACT

Several studies aim to understand the Yield Gaps, measuring possible increasing on crop yields to meet the growing demand for food. These studies usually use complex models to estimate crop yields and yield gaps, mainly presenting values of possible yield increment, without specific management attributes involved in Yield Gap. The present study aims to classify attributes related with crop management involved in soybean crop yield, presenting a multivariate approach and variable importance based classification. In 2021/2022 season, 21 attributes related with crop management, from 470 SLC Agrícola's soybean tillages were evaluated, representing 8 different farms in MA, BA, GO, MS and MT states. Soybean yield ranges from 716.0 to 6044.0 kg ha⁻¹, and comparison of averages shows highest yields in Pantanal, Planalto and Paladino farms, in MS e BA states. High positive correlations were found between soybean yield and base saturation (V%), soil pH, phosphorus (P) and boron (B) and negative correlation between soybean yield and potential acidity (H+Al) and latitude (Y). Four groups were distinguished in unsupervised classifications, two classified with low soybean yields, of which one of them had higher potential yields and total Yield Gaps than other, and others two groups with high soybean yield, separated by differences in clay, fertility and soil nutrients attributes. Random Forest (RF) was used as a soybean yield model, and achieved high adjustment in trial ($R^2 = 0,97$), assuring relative variable importance analysis (IRV). IRV values show V%, H+Al, solar radiation, P and cumulated rainfall as major attributes in soybean yield variation. Volume and variability of data allowed good results identifying attributes of crop management Yield Gaps within proposed analysis flow.

¹ Master Dissertation in Plant Science, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. (80p.) November, 2022.

SUMÁRIO

	Página
1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA	3
2.1 Situação atual do cultivo de soja no Brasil.....	3
2.2 Caracterização da agricultura nas regiões centro-oeste e nordeste do Brasil	5
2.3 Solos do Cerrado Brasileiro	8
2.4 Sistemas de produção no Cerrado	9
2.5 Lacunas (“gaps”) de Produção	11
3. MATERIAL E MÉTODOS	15
3.1 Caracterização do local de execução do estudo	15
3.2 Caracterização das áreas estudadas	19
3.3 Definição da amostragem e dos atributos químicos do solo.....	21
3.4 Dados meteorológicos	24
3.5 Produtividade real.....	25
3.6 Produtividade potencial	25
3.7 Produtividade atingível	27
3.8 “Yield Gap”	27
3.9 Análises estatísticas e modelagem da produtividade de grãos.....	28
4. RESULTADOS E DISCUSSÃO.....	31
4.1 Análise descritiva	31
4.2 Produtividade Potencial e “Yield Gap”	38
4.3 Análise multivariada.....	42

	Página
4.4 Modelagem da produtividade da soja	53
4.5 Importância relativa das variáveis	54
5. CONCLUSÃO.....	58
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	59
7. APÊNDICES	64

LISTA DE TABELAS

	Página
Tabela 1. Caracterização das áreas de estudo quanto a localidade, altitude, percentual médio de argila, classificação de solo, média de precipitação, clima e bioma.	19
Tabela 2. Datas inicial, final e preferencial de semeadura, número de talhões e área total das fazendas estudadas.	20
Tabela 3. Análise descritiva dos atributos meteorológicos, fitotécnicos e de solo coletados nas fazendas de produção.	32
Tabela 4. Comparativo de médias dos atributos meteorológicos, fitotécnicos e de solo entre as fazendas de produção.	34
Tabela 5. Resumo dos componentes principais (CP) obtidos da análise de componentes principais dos atributos da soja, solo e meteorológicos dos locais de cultivo.	45
Tabela 6. Contagem de talhões das fazendas em cada grupo do dendrograma.	49
Tabela 7. Comparativo de médias e análise ANOVA dos atributos meteorológicos, fitotécnicos e de solo entre os grupos separados pela classificação não supervisionada. ...	52
Tabela 8. Métricas de avaliação do modelo de <i>machine learning</i> Random Forest.	54

LISTA DE FIGURAS

	Página
Figura 1. Produtividade média anual de soja no Brasil entre as safras de 1976/1977 e 2021/2022. Adaptado de CONAB, 2022.	4
Figura 2. Área com produção de soja e representatividade de cada região ao longo dos anos. Adaptado de CONAB, 2022.	4
Figura 3. Produtividade média da soja (kg ha ⁻¹) entre as safras de 1976/1977 e 2021/22 na região Centro-oeste (A) e entre as safras 1979/1980 e 2021/22 na região Nordeste (B). Adaptado de CONAB, 2022.	5
Figura 4. Divisão dos estados brasileiros em grandes regiões (IBGE, 2020) ¹	6
Figura 6. Mapa do Brasil com a localização das áreas de estudo nas camadas de (A) Unidades Federativas, (B) média anual de pluviometria (ANA), (C) Classificação climática de Koppen (ALVARES et al., 2013) e (D) Biomas (IBGE, 2019). Identificação das fazendas de produção: 1: Fazenda Parnaíba; 2 - Fazenda Planeste; 3 – Fazenda Paladino; 4 – Fazenda Pamplona; 5 – Fazenda Pantanal; 6 – Fazenda Planalto; 7 – Fazenda Paiaguás; 8 – Fazenda Planorte.	16
Figura 7. Valores diários de precipitação (mm), temperaturas mínima e máxima (°C) e Radiação solar (MJ/m ²) das áreas de estudo, registrados por estações da Zeus [®] instaladas <i>in loco</i>	18
Figura 8. Médias de atributos relacionados às condições químicas e físicas do solo das áreas de estudo. Abreviações: PB: Fazenda Parnaíba; PE - Fazenda Planeste; PD – Fazenda Paladino; PP – Fazenda Pamplona; PT – Fazenda Pantanal; PL – Fazenda Planalto; PG – Fazenda Paiaguás; PR – Fazenda Planorte.	23

Figura 9. Heatmap de coeficientes de correlação de Pearson, obtido a partir de atributos da soja, solo e meteorológicos dos locais de cultivo. * indica correlação significativa (p-valor < 0,05). Abreviações: Prod.: Produtividade; YGt: “Yield gap” total; YGm: “Yield gap” de manejo; SB: Soma de bases; V: Saturação de bases; H+Al: Acidez potencial do solo; CTC: Capacidade de troca de cátions; Ca: Cálcio; Mg: Magnésio; K: Potássio; B: Boro; P: Fósforo; MO: Matéria Orgânica; Zn: Zinco; Mn: Manganês; S: Enxofre; Cu: Cobre; Fe: Ferro; Al: Alumínio; Y: Latitude da lavoura.37

Figura 10. Gráficos de produtividades real, atingível e potencial máximo, destacando os valores de “Yield Gap” total (YGt) e “Yield Gap” de manejo (YGm) de cada fazenda de produção. Abreviações: Prod. Real – Produtividade média obtida na safra 2021/22; Identificação: 1: Fazenda Parnaíba; 2 - Fazenda Planeste; 3 – Fazenda Paladino; 4 – Fazenda Pamplona; 5 – Fazenda Pantanal; 6 – Fazenda Planalto; 7 – Fazenda Paiaguás; 8 – Fazenda Planorte.39

Figura 11. Distribuição das médias de (A) produtividade potencial (kg ha⁻¹), (B) “Yield GAP” total (kg ha⁻¹), (C) produtividade de soja obtida na safra 2021/2022 (kg ha⁻¹) e (D) “Yield GAP” de manejo (kg ha⁻¹) no Brasil.42

Figura 12. Dendrograma com classificação não supervisionada das lavouras demonstrando os grupos (A) e gráfico de distância acumulada demonstrando a altura do corte (B).44

Figura 13. PCA biplots dos componentes principais 1 e 2 (A) e componentes principais 2 e 3 (B), indicando as áreas de estudo (pontos) e os carregamentos de cada atributo (setas). Os pontos próximos correspondem a observações com pontuações semelhantes nos componentes da PCA. O comprimento das setas indica a variação dos atributos no componente, enquanto os pontos entre eles indicam suas correlações. Abreviações: PD – Paladino; PB – Parnaíba; PR – Planorte; PT – Pantanal; PL – Planalto; PE – Planeste; PG – Paiaguás; PP – Pamplona.....47

Figura 14. PCA biplots dos componentes principais 1 e 2 (A) e componentes principais 2 e 3 (B), indicando a classificação de grupos da análise hierárquica (pontos) e os carregamentos de cada atributo (setas). Os pontos próximos correspondem a observações com pontuações semelhantes nos componentes da PCA. O comprimento das setas indica a variação dos atributos no componente, enquanto os pontos entre eles indicam suas correlações.51

Figura 15. Produtividade real estimada pelo modelo Random Forest (RF) em função da produtividade real obtida (observada). Conjunto de dados de treinamento (em verde): 352 observações e; conjunto de dados de teste (em azul): 118 observações. 53

Figura 16. Importância relativa das variáveis avaliadas na redução de produtividade da Soja, agrupados por (A) Grupo de variáveis e aberto nos grupos de (B)Fertilidade do solo, (C) Climáticas, (D) Macronutrientes e (E) Micronutrientes. Abreviações: Fert – Fertilidade do solo; Macr – Macronutrientes do solo; Micr – Micronutrientes do solo. 55

Figura 17. Importância relativa das variáveis avaliadas no modelo de produtividade da Soja, classificadas pela ordem de maior para o menor impacto na produtividade. 57

1. INTRODUÇÃO

A área de cultivo de soja no Brasil vem aumentando constantemente, tendo passado a marca dos 40 milhões de hectares na safra 2021/22. Além do crescimento de área, nota-se também um significativo incremento de produtividade da cultura ao longo dos últimos anos, em decorrência do aumento do potencial genético das cultivares e das melhorias das práticas de manejo desta cultura, como o plantio nas melhores épocas, a adubação e correção do solo ao longo dos anos. Mesmo com este incremento, há necessidade de aumento de produção desta cultura para atendimento à crescente expectativa de demanda futura.

Estudos buscam demonstrar a lacuna de produtividade existente entre as médias obtidas para a cultura e seus potenciais produtivos, demonstrando tanto os espaços de aumento de produção, como também, as causas específicas destas lacunas, a fim de buscar soluções para as suas reduções. A partir dos valores de produtividades potencial e atingível da cultura, é possível obter-se os valores de “Yield Gap” total e de manejo, sendo o “Yield Gap” de manejo aquele que tem por objetivo demonstrar a lacuna possível de ser reduzida a partir de práticas agrícolas. Modelos de simulação de cultivos, como APSIM e DSSAT, costumam ser utilizados para a obtenção das produtividades potencial e atingível, bem como realizam a modelagem dos atributos que as afetam, porém são de uso moroso e requerem estudos de calibrações dos dados (NÓIA JÚNIOR & SENTELHAS, 2020).

A hipótese deste estudo é que o uso de estatística multivariada associada às técnicas de *machine learning*, e ao conceito de “Yield Gap”, possa ser útil para compreender as causas de variação de produtividade da cultura da soja e a obtenção de informações de práticas de manejo associadas, facilitando o processo de análise dos modelos preditivos de produtividade da cultura, a fim de identificar as principais causas das lacunas de produtividade.

O objetivo deste trabalho foi avaliar o quanto a produtividade da soja é impactada pelos atributos de solo e de clima vinculados ao manejo da cultura, propondo processo de avaliação com base em estatística multivariada e de classificação de importância de variáveis.

2. REVISÃO BIBLIOGRÁFICA

2.1 Situação atual do cultivo de soja no Brasil

O agronegócio brasileiro tem uma grande relevância para a economia do Brasil e do mundo, apresentando um total de 27,5 % do PIB nacional no ano de 2021, o equivalente a R\$2,39 trilhões. Somente o ramo agrícola foi responsável por 14,6% do PIB, enquanto o ramo pecuário foi responsável pelos demais 6,8 %. Segundo o último levantamento de 2020, mais de 16,4 milhões de pessoas estavam ocupadas com trabalhos oriundos do agronegócio, representando 20,1 % das ocupações do país (CEPEA, 2020).

A soja é uma das principais commodities agrícolas do Brasil, ocupando, no ano de 2022, mais de 40,92 milhões de hectares, o que representa 47,5 % da área brasileira ocupada com lavouras temporárias neste ano. Em 2021, foram produzidas mais de 134,9 milhões de toneladas de soja no país (IBGE, 2021). Nos últimos três anos, o Brasil foi o maior produtor mundial de soja, seguido pelos Estados Unidos e pela Argentina, ocupando também a posição de maior exportador mundial do grão (USDA, 2022).

A produtividade da cultura da soja no Brasil apresentou um incremento de 73 % entre as safras de 1976/1977 e 2021/22, quando atingiu uma produtividade média de 3026 kg ha⁻¹ (Figura 1).

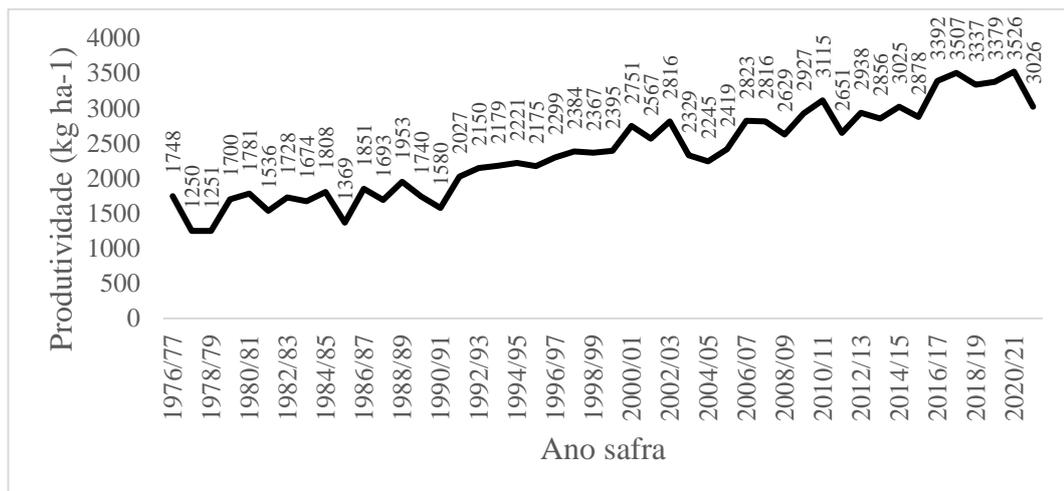


FIGURA 1. Produtividade média anual de soja no Brasil entre as safras de 1976/1977 e 2021/2022. Adaptado de CONAB, 2022.

As regiões Centro-oeste e Nordeste foram as que mais cresceram em representatividade no cultivo de soja entre as safras de 1976/1977 e 2021/2022, saindo de 5 para 46 % e de 0 a 9 %, respectivamente (Figura 2), além de apresentarem crescimento consistente na produtividade ao longo dos anos, com aumento de 149 % entre as safras de 1976/1977 e 2018/2019 na região Centro-oeste e de 213 % entre as safras de 1980/1981 e 2018/2019 na região Nordeste, como observado na figura 3 (CONAB, 2022).

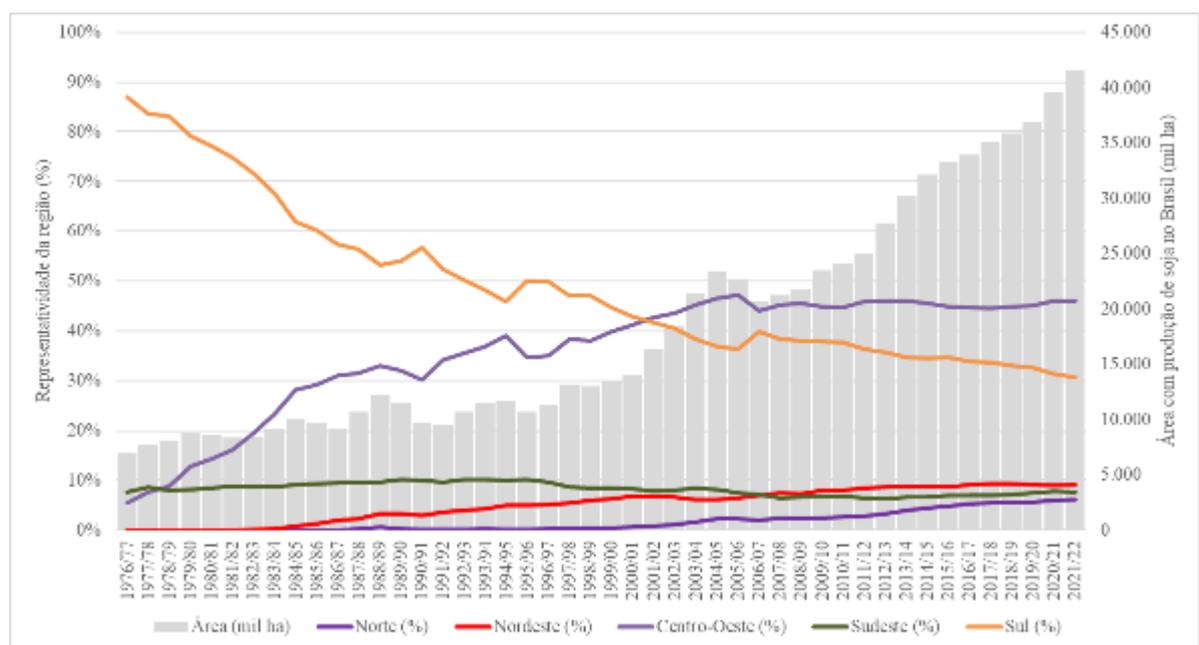


FIGURA 2. Área com produção de soja e representatividade de cada região ao longo dos anos. Adaptado de CONAB, 2022.

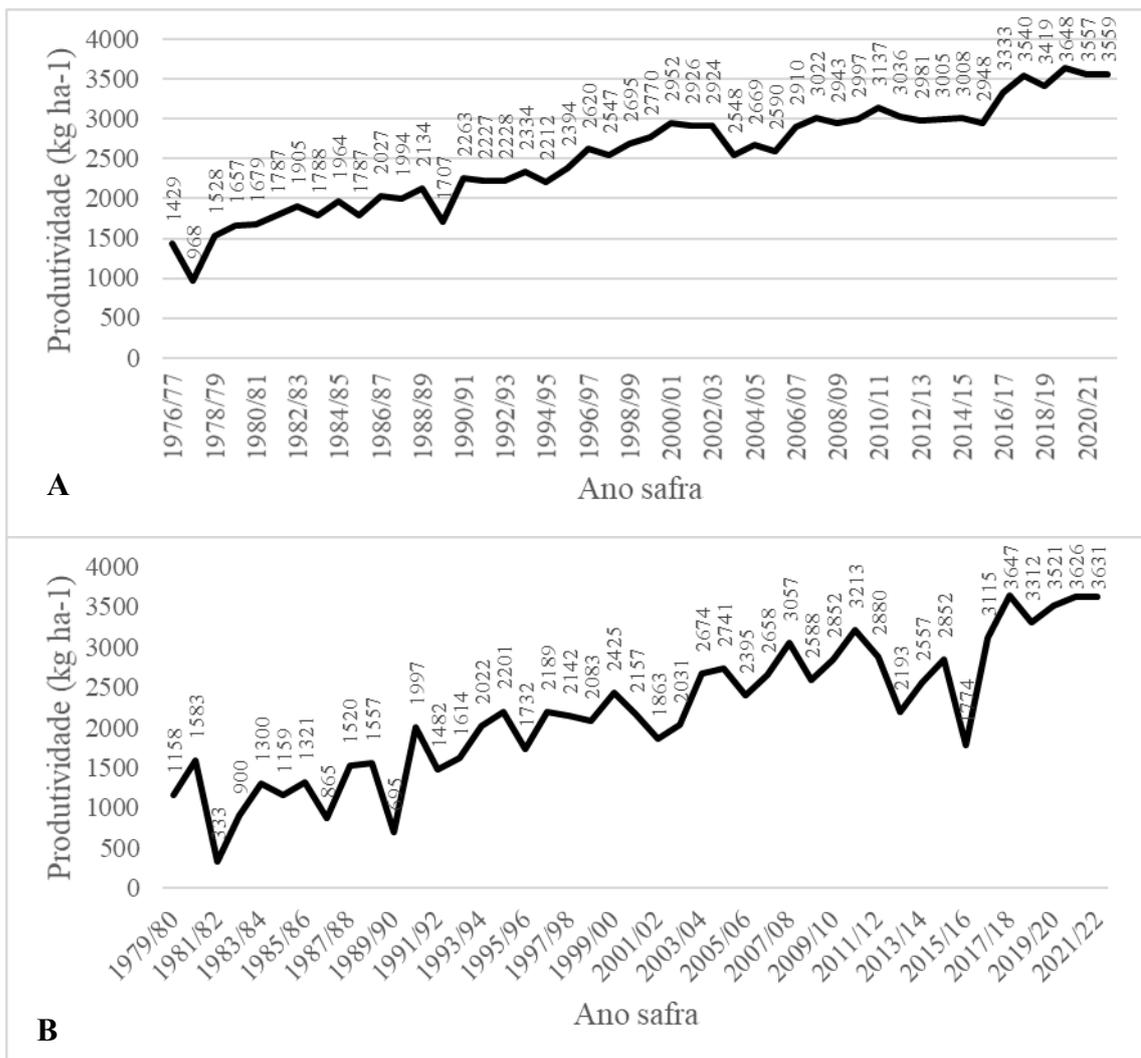


FIGURA 3. Produtividade média da soja (kg ha⁻¹) entre as safras de 1976/1977 e 2021/22 na região Centro-oeste (A) e entre as safras 1979/1980 e 2021/22 na região Nordeste (B). Adaptado de CONAB, 2022.

2.2 Caracterização da agricultura nas regiões centro-oeste e nordeste do Brasil

O Brasil é composto por cinco regiões, Sul, Sudeste, Centro-Oeste, Nordeste e Norte (Figura 4). Cada região brasileira apresenta diferentes potenciais agrícolas para as mais diversas culturas, apresentando diferentes tamanhos de áreas das propriedades rurais existentes.

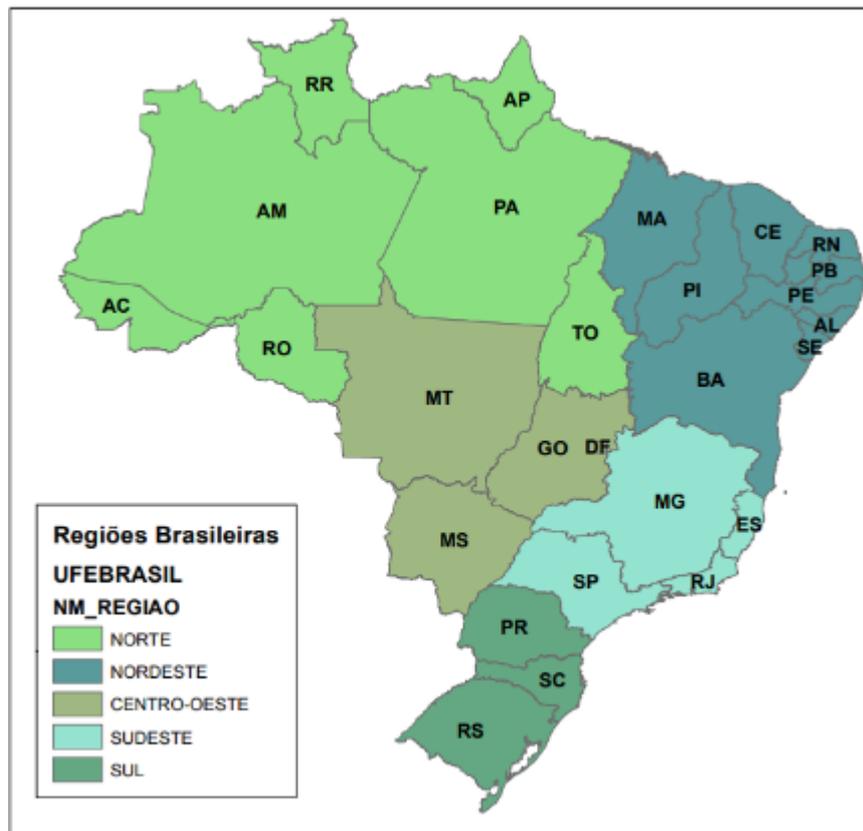


FIGURA 4. Divisão dos estados brasileiros em grandes regiões (IBGE, 2020)¹.

Segundo o último censo agropecuário divulgado pelo IBGE, as lavouras temporárias ocuparam em 2017 uma área de 67,6 milhões de hectares no território brasileiro, com distribuição entre as regiões de 4 % (Norte), 12 % (Nordeste), 17 % (Sudeste), 29 % (Sul) e 39 % (Centro-Oeste). O censo demonstrou que o tamanho médio das lavouras temporárias colhidas nos estabelecimentos agropecuários (área colhida/número de estabelecimentos) foi de 275,5 hectares em 96.574 estabelecimentos no Centro-Oeste, e de 4,75 hectares em 1.644.548 estabelecimentos no Nordeste. Apesar da região nordeste possuir 16 vezes mais estabelecimentos agropecuários, é responsável por uma área total que representa somente 29 % da área da região centro-oeste (IBGE, 2017).

A soja, o milho e a cana-de-açúcar foram as culturas com maior área cultivada no Brasil, representando 45 %, 22 % e 13 % na safra agrícola 2018/19, respectivamente, sendo

a região Centro-Oeste a com maior representatividade em área de produção de soja (45 % da área brasileira) e milho (50 % da área brasileira). A segunda maior em relação à cana-de-açúcar (20 % da área brasileira). Além disso, o Centro-Oeste foi a região com maior percentual de área de produção de sorgo (44 %) e girassol (92 %).

A região Nordeste se destacou no cenário nacional pela representatividade de área das culturas de castanha-de-caju (100 %), melão (90 %), coco-da-baía (82 %), manga (73 %), maracujá (69 %), feijão (46 %) entre outras, demonstrando uma diferença em representatividade das culturas em relação à região Centro-Oeste, que é mais relevante em culturas de grãos. Apesar disso, soja e milho foram as culturas que mais ocuparam área em ambas as regiões, sendo que milho ocupou 30 % e 22 % e a soja 55 % e 31 % das áreas produtivas do Centro-Oeste e Nordeste, respectivamente (IBGE, 2020)².

Esta diferença em relação às culturas nas regiões pode ser explicada por características climáticas. A região Centro-Oeste comporta o cultivo de segunda safra de grãos e fibra devido à precipitação média anual predominante ser na faixa entre 1600 e 2500 mm, sendo que o leste goiano apresenta as menores médias de precipitação, em uma faixa entre 1300 e 1600 mm. O clima desta região é dividido entre duas classes, tropical de monção (Am) e tropical de inverno seco (Aw) pela classificação de Köppen. A região Nordeste tem valores médios de precipitação anual consideravelmente menores, sendo abaixo de 700 mm na região do bioma caatinga até uma faixa entre 1300 e 1600 mm no oeste baiano, semelhante ao leste goiano, apresentando média superior a 1900 mm somente na área litorânea. Nesta região, predominam quatro diferentes climas segundo a classificação de Köppen, tropical de inverno seco (Aw), tropical de verão seco (As), seco de baixa latitude e altitude (BSh) na região da caatinga, e tropical sem estação seca (Af) principalmente na região litorânea da Bahia (ALVARES et al., 2013).

A cultura da soja esteve presente em 1.544 estabelecimentos agropecuários da região Nordeste (0,09 % do total de estabelecimentos da região), com área média de 1.866 hectares, e em 22.385 estabelecimentos agropecuários da região Centro-oeste (23 % do total de estabelecimentos da região), com área média de 641 hectares (IBGE, 2017).

2.3 Solos do Cerrado Brasileiro

As características edafoclimáticas da região do cerrado brasileiro contribuíram com os diversos tipos de solo neste bioma. Uma das principais características é a alta acidez, fator que caracteriza os solos do Cerrado, bem como a predominância de Latossolos (Figueiredo et al., 2009). No geral, as altitudes são consideradas de baixas a médias, com altitudes entre 300 e 800 m, porém chegando a altitudes de 1600 m na região da Serra dos Pirineus. Além disso, estão em uma formação geologicamente antiga e com relevo relativamente acidentado.

Por se tratar de uma região de formação antiga, os solos foram muito influenciados por ações de intemperismos (agentes água e vento). A lixiviação fez com que, ao longo dos anos, a fertilidade tenha reduzido significativamente. A profundidade e porosidade elevadas são características relevantes dos solos que compõem o Cerrado, permitindo uma alta taxa de infiltração da água (Oliveira et al., 2004). Apesar desta alta taxa de infiltração, não há uma grande capacidade dos solos armazenarem esta água.

Seguindo a definição do Sistema Brasileiro de Classificação de Solos (SiBCS), os Latossolos têm como características as argilas dominadas pelos óxidos de ferro, de alumínio, de silício e de titânio, sendo argilas de baixa capacidade de troca de cátions (CTC), e que apresentam baixo pH, além de geralmente apresentarem baixa fertilidade (dos Santos et al., 2018). Suas condições físicas favorecem o uso para a agricultura, sendo bem estruturados e porosos, requerendo, geralmente, a correção da acidez e a adubação elevada.

O Cerrado brasileiro também possui os Argissolos, podendo ser vermelho escuro ou vermelho-amarelo e com grande sensibilidade à erosão. Os Argissolos têm baixo teor de fósforo, potássio, cálcio, magnésio, zinco, boro e cobre, além de apresentarem alta acidez e alta saturação de alumínio. O Argissolo, depois dos Latossolos, é o tipo de maior ocorrência no Brasil e tem como característica uma clara e abrupta separação entre os horizontes diagnósticos no que se refere a cor, estrutura e textura (Camargo et al., 2014). Quando apresentam classificação textural média a argilosa, são adequados para a agricultura, pois apresentam capacidade de retenção de água e boas reservas minerais (Cortez et al., 2015).

De modo geral, os solos do Cerrado apresentam baixa fertilidade, que pode ser resolvido com calcário para correção de acidez, adubação fosfatada e adubação potássica, permitindo assim o desenvolvimento agrícola. Na segunda metade do século XX o uso da terra passou a ser mais intensivo, e as produtividades das culturas na região foram aumentadas, devido e grande parte ao desenvolvimento das técnicas de correção acima mencionadas (Oliveira, 2016).

2.4 Sistemas de produção no Cerrado

O Cerrado brasileiro abrange aproximadamente 24% do território nacional, com 1389 municípios em 12 unidades federativas e no Distrito Federal (IBGE, 2022), e proporciona grande contribuição para a manutenção da biodiversidade e segurança alimentar (DOS REIS et al, 2021).

O uso e cobertura do solo do Bioma Cerrado é dividido nas categorias: floresta, não-floresta e não-vegetação natural; agricultura anual e perene, áreas urbanas, mineração, mosaicos ocupacionais, pastagens plantadas, silvicultura e solo descoberto. Encontrou-se

que o uso antrópico ocupa 43,4% desse bioma, sendo que 8,54% do cerrado era ocupado por culturas anuais de larga escala, como soja, algodão e milho (Scaramuzza, 2017).

Na figura 5 observam-se os principais sistemas de produção de grandes culturas temporárias no cerrado. Devido ao cerrado apresentar sazonalidade de regimes pluviométricos ao longo de seu território, algumas regiões, como o Mato Grosso e o sul do Maranhão, permitem a realização de cultura de segunda safra e outras, como Bahia e Piauí, não. Em regiões onde não há a ocorrência de culturas de segunda safra, as semeaduras de soja e algodão ocorrem a partir do final do mês de outubro e início do mês de dezembro, respectivamente.

Em regiões de período pluviométrico maior, é possível o cultivo em segunda-safra, principalmente com as culturas de algodão e milho, para as quais a colheita de soja deve ocorrer em período mais antecipado, preferencialmente até final de janeiro para o algodão e final de fevereiro para o milho (SLC Agrícola, 2022).

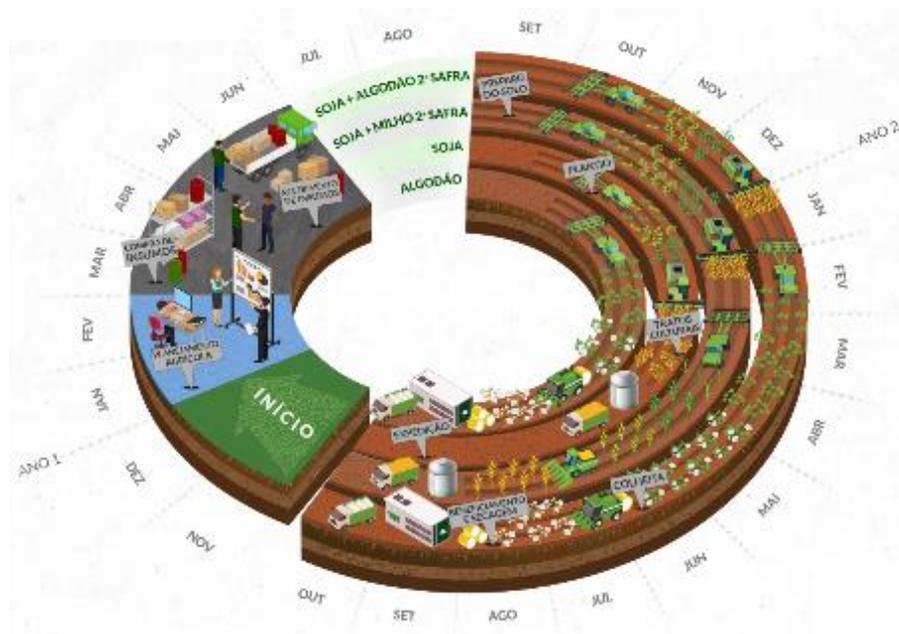


FIGURA 5. Ciclo de produção das culturas de Soja, Milho e Algodão na SLC Agrícola (SLC Agrícola, 2022).

2.5 Lacunas (“gaps”) de Produção

O estudo do “Yield Gap” se mostra importante para a identificação de regiões com potencial de acréscimo produtivo, dada a demanda por uma intensificação sustentável da produção agrícola. São quatro motivos que tornam esta busca bastante relevante. Primeiramente, pela obtenção de informações sobre os fatores que impactam a produtividade dos cultivos atuais, possibilitando a melhoria de práticas para redução do “gap”. Segundo, para guiar os esforços e investimentos de pesquisas, demonstrando os pontos de maior retorno para a segurança alimentar para cada região específica. Terceiro, para possibilitar a avaliação de impactos de mudanças climáticas em cenários futuros sobre o uso de terras e de recursos naturais. Quarto, o entendimento destes “gaps” e como reduzi-los são importantes para modelos econômicos que estimam as demandas e disponibilidades de recursos futuros (ITTERSUM et al, 2013).

O conceito de “yield gap” define a diferença entre a produtividade obtida em um determinado local frente a um valor de referência de produtividade deste mesmo local. Portanto, o “yield gap” representa a redução de produtividade em relação ao valor de referência. (ITTERSUM & CASSMAN, 2013).

O “yield gap” pode ser avaliado em um ambiente produtivo, comparando-se o potencial máximo de produção de todas as culturas em determinada área em um ano agrícola frente ao valor efetivamente produzido, demonstrando o quanto há de espaço para ganho de produção pelo ajuste da sucessão de culturas. Desta forma, o objetivo é entender como otimizar a produtividade da área como um todo, e não de alguma cultura específica (GUILPART et al., 2017). Contudo, o “yield gap” é mais comumente avaliado para cada cultura individualmente, objetivando-se entender como alcançar a maior eficiência sobre

esta cultura específica, partindo-se de valores de referência como produtividade potencial, produtividade atingível e produtividade real.

A produtividade potencial, representada usualmente como Y_p (“potential yield”), é aquela onde são consideradas as plantas em desenvolvimento sem restrições de água e de nutrientes, e sem perdas por estresses bióticos e abióticos, tendo sua produtividade máxima definida por questões ambientais do ambiente de produção, como temperatura, radiação solar e concentração de CO_2 atmosférico. Como é considerada a disponibilidade total de água e de nutrientes, o solo não é importante para a estimativa de produtividade potencial de diferentes regiões. Este valor é obtido por modelos considerando condições ótimas de desenvolvimento ou por ensaios de pesquisa onde as plantas se desenvolvem sem nenhum estresse (SENTELHAS et al., 2016).

A produtividade atingível (Y_a ou “attainable yield”) é definida como o máximo possível de produção considerando as condições climáticas de cada região, de modo que a planta terá disponível a quantidade de água oriunda somente de precipitação e que possa ser acumulada no solo. Os demais fatores são considerados em condições ótimas para o desenvolvimento das plantas. Portanto, a diferença entre Y_p e Y_a é considerada como o “yield gap” por deficiência hídrica, representando o quanto a planta deixou de produzir devido somente a restrições quanto à disponibilidade hídrica (ITTERSUM et al., 2013).

Para a quantificação do “yield gap” de manejo, utiliza-se a diferença entre Y_a e a produtividade real da área de estudo (Y_f ou “farm yield”) - podendo ser de um talhão, de uma propriedade rural inteira, ou de qualquer outro ambiente de produção. O “yield gap” de manejo é influenciado pelas práticas utilizadas pelo produtor, desde os manejos de solo e escolha de genótipo, até a qualidade do controle de estresses bióticos ao longo do período

produtivo. (ITTERSUM et al., 2013; SENTELHAS et al., 2015; FAO and DWFI. 2015; BATTISTI et al. 2018).

São utilizados ainda outros termos, como a produtividade teórica, que considera somente a capacidade biológica da planta produzir, desconsiderando a interação com o ambiente (FAO & DWFI. 2015), além da divisão da produtividade real entre áreas irrigadas e não irrigadas (SENTELHAS et al., 2015; FAO & DWFI., 2015), e também a separação em dois grupos de produtores pelo nível técnico das práticas de manejo, sendo separado entre a produtividade média e a produtividade dos melhores produtores (SENTELHAS et al., 2015).

A obtenção do valor de Y_a para considerar em um estudo local pode ser realizado de pelo menos quatro formas, sendo essas: (1) Utilização de experimentos com a cultura de interesse; (2) Realização de concursos entre produtores locais, estimulando a obtenção de altas produtividades; (3) Questionários respondidos pelos produtores, com a obtenção dos valores máximos de produtividade; (4) Utilização de modelos que simulam o cultivo. No caso dos modelos, são utilizados para estimar a produtividade, através da interação entre o genótipo x ambiente de produção x práticas de manejo (ITTERSUM et al., 2013). A escolha do mais adequado está atrelado à abrangência do estudo e o objetivo final de uso dos resultados. Um fator a ser considerado é a representatividade dos dados pela sua quantidade e confiabilidade.

A partir da definição dos valores de Y_p , Y_a e Y_f , pode-se ainda determinar os valores de eficiência climática e eficiência de manejo. A eficiência climática pode ser definida para cada ambiente de produção, por meio da divisão entre a produtividade atingível e a produtividade potencial. Já a eficiência de manejo é dada pela divisão entre a produtividade média real obtida e a produtividade atingível do ambiente de produção, mostrando quanto cada local produziu em relação ao valor de referência. O valor de eficiência de manejo é

importante para elucidação do tamanho destas lacunas em cada ambiente de produção e diferentes regiões produtoras. Porém, apesar desta quantificação ser possível de ser obtida, cada área específica necessita de um estudo individual para a identificação das causas do “yield gap” de manejo das culturas (BATTISTI et al., 2018).

3. MATERIAL E MÉTODOS

3.1 Caracterização do local de execução do estudo

O estudo foi realizado durante a safra 2021/22, em oito fazendas de produção comercial de soja sequeiro, pertencente à SLC Agrícola. As fazendas são distribuídas nos estados de Mato Grosso (MT), Mato Grosso do Sul (MS), Goiás (GO), Bahia (BA) e Maranhão (MA) (Figura 6A).

Segundo o mapa de isoietas (Figura 6B), a média de precipitação acumulada anual nessas regiões varia de 1100 a 1200 mm, na fazenda Paladino (BA), a 1900 a 2000 mm, na fazenda Paiaguás (MT). Quando definimos o clima das fazendas de produção, seguindo a classificação de Köppen (Alvares et al., 2013), verifica-se que são do tipo tropical úmido (Aw) e monçônico (Am), caracterizados por verão chuvoso e inverno seco (Figura 6C). Outra característica importante é que todas as fazendas de produção estão localizadas no bioma Cerrado (Figura 6D).

A escolha das lavouras para o estudo levou em consideração alguns aspectos, como: diferentes ambientes de produção em relação a teores de argila do solo, regime hídrico e fertilidade do solo, informações históricas disponíveis, principalmente fertilidade do solo, histórico de cultivo, produtividade e conhecimento das áreas pelos profissionais envolvidos na gestão das fazendas. Nessas áreas, adotam-se o sistema de semeadura direta, utilizando-

nas fazendas de produção Planorte, Paiaguás e Planalto, apresentando consistência de chuvas a partir do mês de setembro. Para as demais fazendas, o período de chuvas mais consistentes iniciou em outubro. Além disso, nota-se que para as fazendas Parnaíba e Planeste, as precipitações se mantiveram até o final do mês de abril, demonstrando um período mais prolongado que o das demais fazendas de produção. As fazendas de produção Pamplona e Paladino foram as que tiveram o início das precipitações mais tarde entre as fazendas avaliadas.

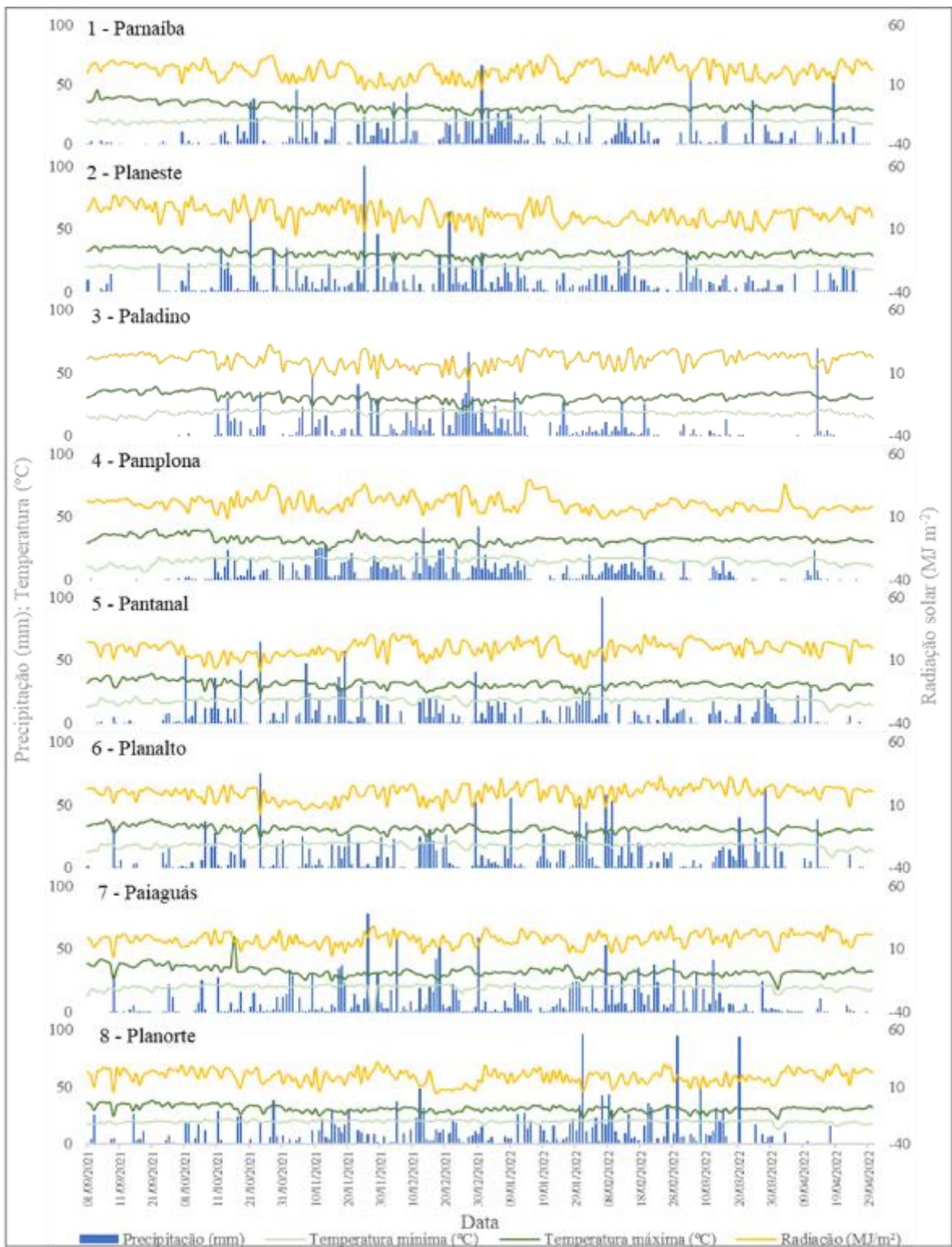


FIGURA 6. Valores diários de precipitação (mm), temperaturas mínima e máxima (°C) e Radiação solar (MJ/m²) das áreas de estudo, registrados por estações da Zeus[®] instaladas *in loco*.

Os solos dos locais das fazendas estudadas são classificados como: Latossolo Vermelho Distrófico – LVd e Latossolo Amarelo Distrófico – LAd, de acordo com o Sistema Brasileiro de Classificação de Solo (Embrapa, 2018). A granulometria dos solos dos locais consta na Tabela 1, e foi determinada pelo método da pipeta, de acordo com procedimentos metodológicos propostos por Embrapa (2017).

TABELA 1. Caracterização das áreas de estudo quanto a localidade, altitude, percentual médio de argila, classificação de solo, média de precipitação, clima e bioma.

Fazenda	Município/UF	Altitude	Argila	Solo ¹	Precipitação ²	Clima ³
		(m)	(%)			
Parnaíba	Tasso Fragoso – MA	560	37	LAd	1400 - 1500	Aw
Planeste	Balsas – MA	570	59	LAd	1500 - 1600	Aw
Paladino	São Desidério - BA	790	18	LAd	1100 - 1200	Aw
Pamplona	Cristalina – GO	990	59	LVd	1400 - 1500	Aw
Pantanal	Chapadão do Sul - MS	820	52	LVd	1400 - 1500	Am
Planalto	Costa Rica – MS	850	65	LVd	1500 - 1600	Am
Paiaguás	Diamantino - MT	600	53	LVd	1900 - 2000	Aw
Planorte	Sapezal - MT	650	61	LVd	1800 - 1900	Aw

MA = Maranhão; BA = Bahia; GO = Goiás; MS = Mato Grosso do Sul; MT = Mato Grosso; LAd = Latossolo Amarelo distrófico; LVd = Latossolo Vermelho distrófico; Aw = Tropical de inverno seco; Am = Tropical de monções

¹Mapa de solos do Brasil (EMBRAPA)

²Precipitação média anual 1977 – 2006 (Agência Nacional de Águas - ANA)

³Classificação climática de Koppen

3.2 Caracterização das áreas estudadas

Ao todo, 470 talhões foram selecionados para estudo, sendo 52 na Fazenda Parnaíba, 76 na Fazenda Planeste, 48 na fazenda Paladino, 52 na fazenda Pamplona, 68 na fazenda Pantanal, 10 na fazenda Planalto, 34 na fazenda Planorte e 130 na fazenda Paiaguás, totalizando uma área de 118.995 hectares (Tabela 2.). A diferença entre as áreas de estudo consiste basicamente na geolocalização (latitude) e características sazonais de cada Fazenda.

Nesse sentido, apesar de compreendidas em um mesmo bioma, as condições climáticas (em termos de amplitude da estação chuvosa, volume pluviométrico e radiação)

inferem diretamente sobre o ambiente de produção da soja e as rotações de culturas que compõem a segunda safra. Tais condições possibilitaram o estudo da produção de soja em sequeiro em diversos sistemas de produção.

TABELA 2. Datas inicial, final e preferencial de semeadura, número de talhões e área total das fazendas estudadas.

	PB	PE	PD	PP	PT	PL	PG	PR	Total
Número de talhões	52	76	48	52	68	10	130	34	470
Área total (mil ha)	17,5	24,3	8,3	6,1	17,6	3,1	29,7	12,1	118,9
Início da semeadura	19/out	01/out	23/out	13/out	04/out	07/out	27/set	29/set	27/set
Final da semeadura	18/nov	07/dez	01/dez	20/dez	29/out	28/out	26/out	25/out	20/dez
Data preferencial de semeadura	25/out	25/out	01/nov	15/out	05/out	05/out	10/out	10/out	-

PB – Parnaíba; PE – Planeste; PD – Paladino; PP – Pamplona; PT – Pantanal; PL – Planalto; PG – Paiaguás; PR - Planorte
set – setembro; out – outubro; nov – novembro; dez – dezembro

Na tabela 2 observa-se a data preferencial de semeadura fornecido pelas fazendas, que são conhecidas a partir de dados de pesquisa internos que correlacionam a produtividade com a semana de plantio ao longo de mais de 5 anos. No período que compreende o final do mês de setembro e início do mês de dezembro, foi semeada a soja nas respectivas lavouras (Tabela 2), com o espaçamento de 0,45 m entre linhas. A adubação de base foi realizada de acordo com o planejamento agrícola das fazendas, considerando a disponibilidade e adequação dos níveis de Fósforo (P) e Potássio (K) no solo e a necessidade da cultura. Para a adubação foram utilizados em média 150 kg ha⁻¹ do fertilizante formulado 00-20-00 (fonte superfosfato simples 20% de P₂O₅) e 130 kg ha⁻¹ de KCl (60% de K₂O), visando uma produtividade de grãos média esperada de 3600 a 4680 kg ha⁻¹, dependendo a fazenda. Os demais tratamentos culturais seguiram as recomendações para cultivo da soja em sequeiro para as respectivas regiões de produção.

3.3 Definição da amostragem e dos atributos químicos do solo

A fertilidade das lavouras foi determinada a partir da média dos resultados de amostragem de solo em grades amostrais com 1 ponto a cada 10 ha e na profundidade de 0,00 – 0,20 m, e todas as análises foram realizadas nos laboratórios IBRA e Solos & Plantas. Para determinar os teores lábeis de P, K, Ca e Mg, foi utilizado o método de resina de extração de íons (van Raij et al., 2001).

O teor de P foi determinado por colorimetria (Murphy e Riley, 1962), usando um espectrômetro UV-VIS. Os teores de Ca e Mg foram determinados por espectrofotometria de absorção atômica, e o Al foi extraído com solução de KCl 1 mol L⁻¹ e titulado com solução de NaOH 0,025 mol L⁻¹. O S foi extraído com 0,01 mol L⁻¹ de Ca(H₂PO₄)₂ (Van Raij et al., 2001) e seu conteúdo foi determinado via ICP-OES.

O teor de matéria orgânica (MO) foi determinado por oxidação por 0,167 mol L⁻¹ de dicromato de potássio e posterior titulação com sulfato ferroso (Walkley e Black, 1934). A acidez potencial (H+Al) foi extraída com solução de acetato de cálcio 1 mol L⁻¹ com pH tamponado a 7,0 e determinada por titulação com 0,025 mol L⁻¹ de NaOH. O pH do solo foi medido usando uma solução de CaCl₂ 0,01 mol L⁻¹ na proporção de 1:2,5 de solo e solução (van Raij et al., 2001). Os micronutrientes Cu, Fe, Mn e Zn foram extraídos com solução tamponada de DTPA (0,005 mol L⁻¹ DTPA + 0,1 mol L⁻¹ de trietanolamina e 0,01 mol L⁻¹ de CaCl₂) a um pH de 7,3, enquanto o B foi extraído em água quente. Posteriormente, os teores de todos os micronutrientes foram determinados via ICP-OES (Iatrou et al., 2015; Van Raij et al., 2001).

Os valores médios de teor de argila, CTC, matéria orgânica, saturação de bases, cálcio, magnésio, fósforo, potássio e enxofre das fazendas constam na figura 8. Destacam-se os teores de argila das fazendas de produção Planorte, Planalto, Planeste e Pamplona, na

faixa dos 60%, e da fazenda Paladino, abaixo dos 20%. As fazendas de produção Planalto e Pantanal apresentam os maiores teores de Matéria Orgânica, enquanto o menor valor é observado na fazenda Paladino. Valores de saturação de bases (V%) acima de 60% são encontrados nas fazendas Paladino, Pamplona, Pantanal, Planalto e Planorte, e destaca-se a fazenda Paiaguás com V% abaixo de 40%.

A fazenda de produção Pamplona destaca-se pelo teor de fósforo (P) acima do limiar de 60 mg dm^{-3} . As fazendas Paladino, Planalto e Pantanal apresentam teores de P acima de 40 mg dm^{-3} , enquanto as demais fazendas apresentam teores médios entre 20 e 40 mg dm^{-3} . O teor médio de potássio (K) mais baixo é encontrado na fazenda de produção Parnaíba, entre 0,10 e $0,20 \text{ cmol}_c \text{ dm}^{-3}$, enquanto as fazendas Pantanal, Planalto e Pamplona apresentam valores acima de $0,30 \text{ cmol}_c \text{ dm}^{-3}$.

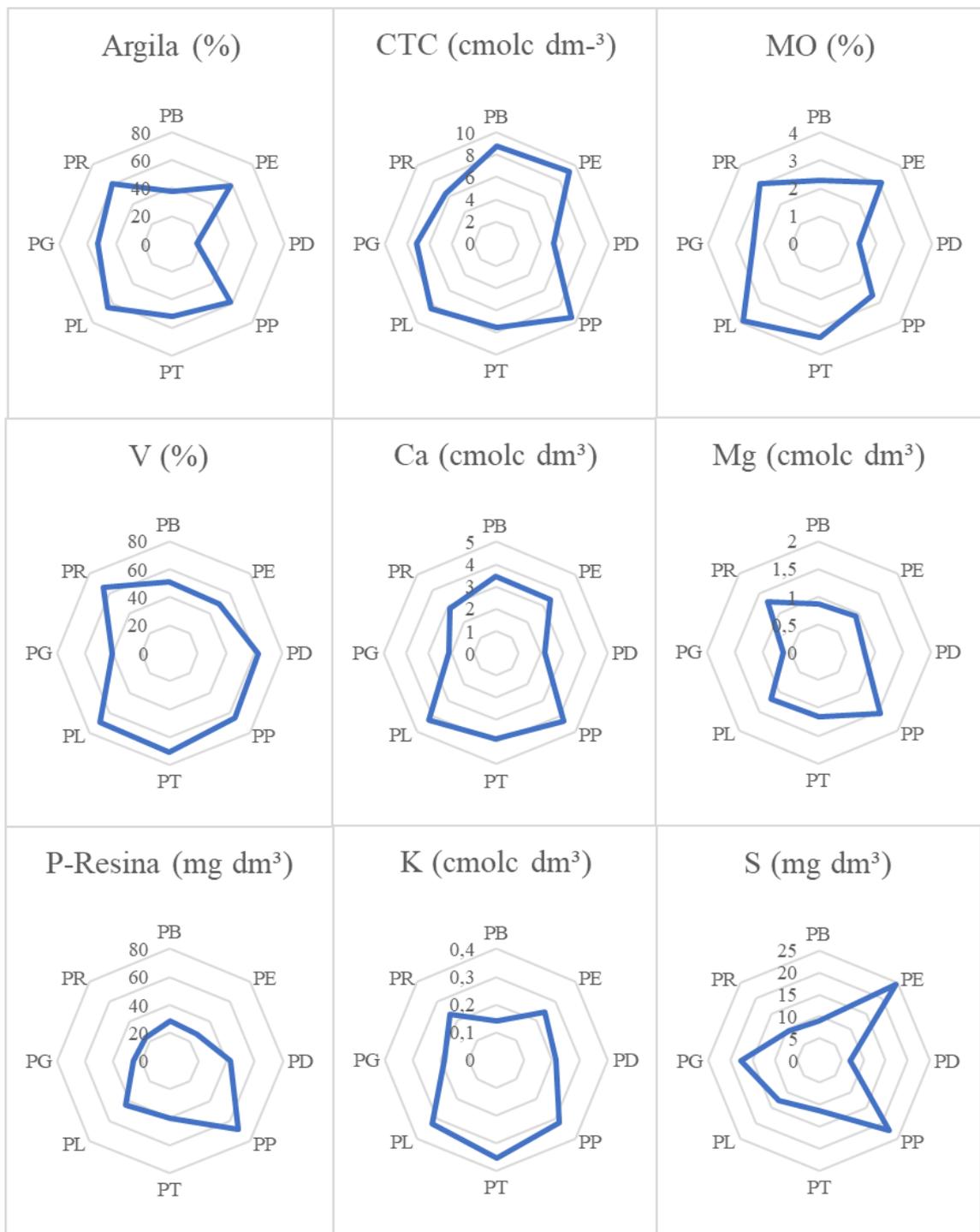


FIGURA 7. Médias de atributos relacionados às condições químicas e físicas do solo das áreas de estudo. Abreviações: PB: Fazenda Parnaíba; PE - Fazenda Planeste; PD – Fazenda Paladino; PP – Fazenda Pamplona; PT – Fazenda Pantanal; PL – Fazenda Planalto; PG – Fazenda Paiaguás; PR – Fazenda Planorte.

3.4 Dados meteorológicos

Os dados meteorológicos utilizados no estudo foram obtidos de estações meteorológicas Zeus®, instaladas em cada uma das fazendas de produção. Os dados coletados foram: precipitação diária, temperatura máxima e mínima do ar, umidade relativa do ar, irradiância solar global e evapotranspiração. Dados de precipitação e temperatura média do ar também foram coletados de pluviômetros digitais alocados ao longo das fazendas de produção (1 pluviômetro representando para cada grupo de 4 – 6 lavouras). Esse procedimento foi adotado devido a disponibilidade dos equipamentos nas fazendas e a variabilidade espacial da pluviometria.

A irradiância solar global extraterreste (Q_0) e o fotoperíodo (N), foram estimados pelas equações 1 e 5, respectivamente:

$$Q_0 = 37,6 \left(\frac{d}{D}\right)^2 \left[\text{sen}Y \cdot \text{sen}\delta \cdot \text{hn} \left(\frac{\pi}{180}\right) + \text{cos}Y \cdot \text{cos}\delta \cdot \text{senhn} \right] \dots\dots\dots(1)$$

$$\delta = 23,45 \cdot \text{sen} \left[\frac{360}{365} \cdot (NDA - 80) \right] \dots\dots\dots(2)$$

$$\text{hn} = \arccos(-\text{tg}Y \cdot \text{tg}\delta) \dots\dots\dots(3)$$

$$\left(\frac{d}{D}\right)^2 = 1 + 0,033 \cdot \cos \left(\frac{360}{365} \cdot NDA \right) \dots\dots\dots(4)$$

$$N = \frac{2 \cdot \text{hn}}{15} \dots\dots\dots(5)$$

em que: $\frac{d}{D}$ é a distância relativa Terra e Sol, Y é a latitude local, δ é a inclinação solar, hn é o ângulo horário do nascer do sol e NDA o dia juliano, compreendido entre 1 e 366.

A estimativa da irradiância solar global incidente foi calculada pela equação de Hargreaves & Samani (1982) (Equação 6), que utiliza como variáveis de entrada a temperatura máxima e mínima do ar.

$$Q_g = 0,16. (T_{max} - T_{min})^{0,5} \dots\dots\dots(6)$$

em que: Q_g é a irradiância solar global incidente ($MJ\ m^{-2}\ dia^{-1}$); T_{max} é a temperatura máxima do ar ($^{\circ}C$) e T_{min} é a temperatura mínima do ar ($^{\circ}C$).

Para estimativa do saldo de radiação do gramado ($MJ\ m^{-2}\ dia^{-1}$), empregou-se a relação linear com a irradiância solar incidente (Q_g , $MJ\ m^{-2}\ dia^{-1}$), obtida por Silva et al. (2007), com R^2 de 0,85.

$$R_n = 0,5498. Q_g \dots\dots\dots(7)$$

em que: R_n corresponde à radiação do gramado.

3.5 Produtividade real

A produtividade de soja (Prod. real) foi determinada por pesagem dos grãos colhidos por processo de colheita mecanizada das lavouras. Foi estabelecida como área útil de colheita os limites das lavouras selecionadas para estudo. Posteriormente, foram aplicados os descontos de impurezas nas unidades de armazenamento e beneficiamento das fazendas e a produtividade extrapolada para uma área de um hectare, corrigida para uma condição de umidade de 14%.

3.6 Produtividade potencial

Para estimativa da produtividade potencial (Prod. potencial) utilizaram-se dados climáticos obtidos em cada local a partir de estações meteorológicas da empresa Zeus Agrotech® aplicados ao modelo da Zona Agroecológica – FAO, que utiliza as variáveis de entrada Q_0 , temperatura do ar, n , N , índice de área foliar (IAF max), taxa de respiração (R),

índice de colheita (C), umidade da parte aérea colhida (U%) e ciclo da cultura (ND), de acordo com a Equação 8:

$$Prod. potencial = \frac{PPB_p \cdot IAFmax \cdot C \cdot ND}{1 - 0,01 \cdot U\%} \dots\dots\dots(8)$$

em que: PPB_p corresponde à produtividade bruta de matéria seca para cultura padrão com IAFmax de 5, em $kg MS ha^{-1} dia^{-1}$.

A PPB_p foi estimada para o período do dia de céu limpo (PPB_c) e de céu nublado (PPB_n), a partir da Equação 9 e 10, respectivamente, em que a relação da parte do dia nublado e claro é dado pelos valores de insolação e fotoperíodo.

$$PPB_c = (107,2 + 8,604 \cdot Q_o) \cdot cTc \cdot \frac{n}{N} \dots\dots\dots(9)$$

$$PPB_n = (31,7 + 5,234 \cdot Q_o) \cdot cTn \cdot \left(1 - \frac{n}{N}\right) \dots\dots\dots(10)$$

em que: cTc e cTn correspondem a um índice de correção da taxa fotossintética em função da temperatura do ar.

A cultura soja, planta C3 de verão, pelo proposto dos autores do modelo pertence ao grupo II e tem seus valores de cTc e cTn determinados pelas equações 11 e 12, de acordo com a temperatura do ar: *se $T \geq 16,5^\circ C$ & $T \leq 37^\circ C$, então:*

$$cTn = 0,583 + 0,014 \cdot T + 0,0013 \cdot T^2 - 0,000037 \cdot T^3 \dots\dots\dots(11)$$

$$cTc = -0,0425 + 0,035 \cdot T + 0,00325 \cdot T^2 - 0,0000925 \cdot T^3 \dots\dots\dots(12)$$

se $T \leq 16,5^\circ C$ ou $T \geq 37^\circ C$, então:

$$cTn = -0,0425 + 0,035 \cdot T + 0,00325 \cdot T^2 - 0,0000925 \cdot T^3 \dots\dots\dots(13)$$

$$cTc = -1,085 + 0,07 \cdot T + 0,0065 \cdot T^2 - 0,000185 \cdot T^3 \dots\dots\dots(14)$$

em que: T é a temperatura média do ar ($^\circ C$).

O somatório diário entre as variáveis PPB_n e PPB_c , resulta no PPB_p diário.

A estimativa de IAFmax para cada cultivar em função da duração do período vegetativo, é obtida de acordo com a Equação 15. Os limites superior e inferior (6,5 e 2,0, respectivamente) dessa variável foram definidos com base no trabalho de Rodrigues et al. (2006).

$$\text{IAFmax} = 0,0851 \cdot \text{PV} + 0,6598 \dots\dots\dots(15)$$

em que: PV é a duração do período vegetativo (em dias) da cultivar de soja, que compreende desde a emergência (VE) até o início da floração (R1).

3.7 Produtividade atingível

Com o intuito de obtenção de produtividade atingível coerente com cada local de estudo, buscou-se no histórico dos últimos 5 anos de cada fazenda de produção (2017/2018, 2018/2019, 2019/2020, 2020/2021 e 2021/2022) a maior produtividade da cultura da soja em áreas de sequeiro. Nesse sentido, a produtividade atingível remete a condições de manejo de solo, fitossanitário e cultural reais de cada unidade produtiva, de acordo com a realidade local.

3.8 “Yield Gap”

Foram utilizados dois valores de “Yield Gap”, sendo eles o “Yield Gap” total (YGt), representando a lacuna total de produtividade, e o “Yield Gap” de manejo (YGM), representando a lacuna de produtividade associada a práticas agronômicas. O YGt foi determinado por talhão pela diferença entre a produtividade real de cada talhão e a máxima produtividade potencial da fazenda. O YGM foi também determinado por talhão pela diferença entre os valores de produtividade real de cada talhão e de produtividade atingível da fazenda.

3.9 Análises estatísticas e modelagem da produtividade de grãos

Por meio do software R (R Development Core Team, 2022), os dados foram submetidos a análise descritiva, através dos cálculos de média, mediana, valores mínimos e máximos, desvio-padrão, coeficientes de variação, curtose e assimetria. Também se procedeu com o teste de normalidade de Shapiro-Wilk ($p \leq 0,05$). Os dados foram submetidos à ANOVA para o comparativo de média entre as fazendas de produção.

A análise de correlação de Pearson ($p \leq 0,05$) foi realizada para verificar as variáveis dependentes que estavam diretamente correlacionadas. O pacote “corrplot” foi acessado para criar o mapa de calor (heatmap), utilizando as funções “color” e “cor.mtest” para criar matrizes de coeficientes e p-valor, respectivamente. Para facilitar a visualização de correlações significativas, foram inseridos asteriscos nas células do mapa de calor.

Com intuito de verificar a formação de grupos por dissimilaridade, realizou-se a análise multivariada de agrupamento (clusters). Para isso, os dados foram padronizados para uma média de 0 e uma variância de 1. O pacote “hclust” foi acessado para construção do dendrograma, utilizando o método de Ward como algoritmo de agrupamento. Em seguida a distância euclidiana foi calculada para construção de uma matriz de dissimilaridade comunitária. A distância euclidiana é a raiz quadrada da soma das diferenças quadradas entre duas variáveis, conforme descrita na Equação 17:

$$dist_{(x,y)} = \sqrt{\sum_i^n (x_i - y_i)^2} \dots\dots\dots(17)$$

O número de grupos foi definido quando a soma acumulada de distância euclidiana correspondeu à 67% do somatório total. Com isso, foram definidos quatro grupos na análise. Os dados dos grupos foram também submetidos à ANOVA.

A relação multivariada entre os atributos do solo, clima e planta foi analisada usando o pacote ‘‘factoextra’’. Utilizou-se os dados padronizados pelo z score e, em seguida, foram submetidos à análise de componentes principais (PCA). Componentes principais (PCs) com autovalores 1 e uma variância explicada de pelo menos 5% da variação total do conjunto de dados foram examinados. A classificação dos autovetores, que são valores que representam o peso de cada atributo em cada componente, foi realizada de acordo com os seguintes critérios: valor absoluto $< 0,30$ indica atributo insignificante; $0,30 - 0,50$ classifica o atributo como moderadamente significativo; e $> 0,50$ classifica o atributo como altamente significativo.

Todos os atributos, exceto Fe e YGm, apresentaram autovetores com peso maior ou igual a $0,50$ – altamente significativo em alguns dos seis PCs, portanto, estes foram retidos na análise. Para uma melhor visualização dos três primeiros CPs extraídos, foram criados gráficos biplots mostrando a carga de cada atributo (setas) e os escores de cada grupo/fazenda (pontos).

A produtividade de grãos de soja obtida nas áreas de estudo foi modelada utilizando as variáveis de solo e clima. Com intuito de entender a importância relativa das variáveis na estimativa da produtividade de grãos, utilizou-se o modelo Random Forest (RF). O RF é um método não paramétrico, que consiste em muitos modelos de árvores individuais treinados a partir de amostras bootstrap dos dados (Breiman, 2001). Os resultados de todas as árvores individuais são agregados para fazer uma única previsão. Esse método também pode classificar a importância relativa da variável preditora com base no erro de previsão de regressão das previsões ‘‘out-of-bag’’ (OOB). RF permuta cada variável e calcula a importância como a mudança resultante no erro OOB (Taghizadeh-Mehrjardi et al., 2016).

Dois parâmetros importantes no método RF são o número de árvores (ntree) e o número de variáveis disponíveis para seleção em cada divisão (mtry) (Houborg e McCabe, 2018).

Random Forest trata problemas comumente encontrados em dados de treinamento, como por exemplo a colinearidade, e captura comportamentos não-lineares entre variáveis, construindo uma infinidade de árvores de decisão. Nesse sentido, 352 lavouras foram separadas para treinamento (75% dos dados) e 118 lavouras para teste (25% dos dados) do modelo. As avaliações de desempenho do modelo foram realizadas pelos cálculos de raiz quadrada do erro médio (RMSE), erro médio absoluto (MAE) e coeficiente de determinação (R^2). Gráficos de validação cruzada foram construídos para todos os conjuntos de dados, para avaliação da produtividade observada \times predita pelo modelo. Todas as análises estatísticas foram realizadas em ambiente R versão 4.1.2 (R Core Team, 2021), utilizando os pacotes “sf”, “raster”, “caret”, “rgdal” e “gstat”.

$$R^2 = 1 - SS_{error}/SS_{total} \dots\dots\dots(18)$$

$$RMSE = \sqrt{1/n \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots\dots\dots(19)$$

$$MAE = 1/n \sum_{i=1}^n |y_i - \hat{y}_i| \dots\dots\dots(20)$$

em que: SS_{error} é soma de quadrados dos resíduos; SS_{total} é soma de quadrados total; n é o número de observações; y_i valor observado da produtividade; \hat{y}_i é o valor estimado da produtividade.

4. RESULTADOS E DISCUSSÃO

4.1 Análise descritiva

Por meio da análise descritiva, da produtividade real da soja na safra 2021/22 observaram-se produtividades mínima de 716,0 kg ha⁻¹, máxima de 6044,0 kg ha⁻¹ e média de 4099,0 kg ha⁻¹ (Tabela 3), superando em 35,4% a média de produtividade nacional de 3026,0 kg ha⁻¹ na safra 2021/22 (CONAB, 2022). O coeficiente de variação encontrado foi de 15,7 %, e o valor positivo de assimetria demonstra que há uma maior frequência de valores acima da média. Os valores da produtividade potencial mínimo, médio e máximo foram de 4674,0, 6244,0 e 7690,0 kg ha⁻¹ respectivamente, com coeficiente de variação de 9,1%.

Avaliando os dados de todas as fazendas, os valores de “Yield Gap” total, obtidos a partir da produtividade potencial e da produtividade real, variaram entre 300,0 kg ha⁻¹ e 5777,0 kg ha⁻¹, com valor médio de 2125,0 kg ha⁻¹ e coeficiente de variação de 39,1% e um coeficiente de assimetria positivo, demonstrando maior frequência de valores de “Yield Gap” total acima da média geral. Os valores mínimo e máximo de “Yield Gap” de manejo foram de 415,0 kg ha⁻¹ e de 5772,0 kg ha⁻¹ respectivamente, com média de 2010,0 kg ha⁻¹.

A radiação acumulada dos talhões variou entre 1690 e 2870 MJ m⁻², com média de 2231 MJ m⁻², e a precipitação acumulada variou entre 669 e 1398 mm m⁻², com média de 1059 mm m⁻².

TABELA 3. Análise descritiva dos atributos meteorológicos, fitotécnicos e de solo coletados nas fazendas de produção.

Atributos	Mínimo	Máximo	Média	Mediana	DP ¹	Coeficientes			S-W ³
						Varição ²	Curtose	Assimetria	
<i>Meteorológicos</i>									
Precipitação (mm)	404	1398	1048	1050	153	14,6	-0,42	1,50	0,950*
Radiação (MJ m ⁻²)	1607	2870	2231	2262	272	12,2	-0,07	-0,65	0,986*
Época de plantio (dias)	-24,0	83,0	8,6	6,0	15,5	181,0	1,48	3,62	0,896*
<i>Fitotécnicos</i>									
Prod. potencial (kg ha ⁻¹)	4674	7690	6224	6272	566	9,1	-0,16	-0,18	0,991*
Prod. real (kg ha ⁻¹)	716	6044	4099	4044	644	15,7	-0,46	2,36	0,967*
YG total (kg ha ⁻¹)	300	5777	2125	2059	831	39,1	0,38	0,26	0,986*
YG manejo (kg ha ⁻¹)	415	5772	2010	1983	625	31,1	0,87	3,85	0,960 ^{ns}
<i>Solo</i>									
pH CaCl ₂	4,50	6,40	5,25	5,10	0,40	7,6	0,58	-0,44	0,952*
H+Al (cmol _c dm ⁻³)	1,00	7,00	3,40	3,50	1,27	37,5	0,16	-0,75	0,973*
Al (cmol _c dm ⁻³)	0,00	0,30	0,04	0,00	0,06	166,7	1,76	3,11	0,625*
Ca (cmol _c dm ⁻³)	1,00	6,90	3,20	3,20	0,89	27,9	0,36	0,67	0,987*
Mg (cmol _c dm ⁻³)	0,30	4,10	1,02	0,90	0,39	38,0	2,07	10,60	0,869*
K (cmol _c dm ⁻³)	0,10	0,90	0,23	0,20	0,12	49,8	1,05	2,10	0,866*
P (mg dm ⁻³)	12,00	127,20	35,56	33,35	14,30	40,2	1,36	3,83	0,919*
S (mg dm ⁻³)	1,40	282,30	14,16	11,30	17,56	124,0	10,05	132,51	0,361*
B (mg dm ⁻³)	0,10	2,40	0,85	0,80	0,37	43,4	0,77	1,61	0,959*
Fe (mg dm ⁻³)	9,80	120,00	36,25	30,30	19,27	53,2	1,97	4,07	0,791*
Mn (mg dm ⁻³)	0,50	25,10	3,27	3,00	1,75	53,5	4,57	50,63	0,763*
Cu (mg dm ⁻³)	0,10	5,00	1,31	1,25	0,69	52,5	1,13	2,42	0,935*
Zn (mg dm ⁻³)	0,20	29,20	2,86	2,40	2,28	79,9	5,34	44,27	0,556*
SB (cmol _c dm ⁻³)	1,40	11,00	4,45	4,40	1,25	28,1	0,65	1,57	0,977*
CTC (cmol _c dm ⁻³)	3,90	12,90	7,84	8,00	1,66	21,2	-0,11	0,28	0,986*
V (%)	24,20	88,60	57,23	56,95	12,16	21,2	0,00	-0,74	0,988*
MO (dag kg ⁻¹)	0,90	5,10	2,68	2,70	0,82	30,8	0,05	-0,21	0,988*
Argila (dag kg ⁻¹)	12,60	83,00	50,02	54,25	18,36	36,7	-0,37	-0,97	0,950*

¹ DP: desvio padrão; ² Variação: coeficiente de variação – CV (%); ³ S-W: valor de W originado do teste normalidade de Shapiro-Wilk, * significativo e ^{ns} não significativo à 0,05 de probabilidade. Época de plantio: diferença em dias entre a data de plantio e a data preferencial de plantio.

Dentre todos os talhões avaliados, os teores de argila variaram entre 12,6 e 83 %, com média de 50 % e um coeficiente de variação de 36,7 % e coeficiente de assimetria negativo, demonstrando maior frequência de valores abaixo da média geral dos talhões.

Os valores médios de saturação de bases, soma de bases, CTC e H+Al para as fazendas de produção foram de 57,3 %, 4,45 $\text{cmol}_c \text{ dm}^{-3}$, 7,84 $\text{cmol}_c \text{ dm}^{-3}$ e 3,4 $\text{cmol}_c \text{ dm}^{-3}$, com coeficientes de variação (CV%) de 21,2 %, 28,1 %, 21,2 % e 37,5 %, respectivamente.

Dentre os macronutrientes, os teores de enxofre, potássio e fósforo apresentaram os maiores coeficientes de variação (CV) (53, 49,8 e 40,2 %, respectivamente). Os teores mínimo e máximo de potássio foram de 0,10 e 0,90 $\text{cmol}_c \text{ dm}^{-3}$ enquanto para fósforo, o mínimo encontrado foi de 12 mg dm^{-3} e o máximo de 127,2 mg dm^{-3} .

O teste de Shapiro-Wilk confirmou a não normalidade dos parâmetros avaliados, com exceção do “Yield Gap” de manejo. Apesar de a normalidade não ser requisito para as análises multivariadas, os outliers foram desconsiderados, mantendo-se somente valores dentro de 3 vezes o intervalo interquartilico para cima e para baixo.

Para melhor entendimento dos ambientes estudados, foi realizado o comparativo de médias dos atributos entre as fazendas de produção (Tabela 4). Observam-se as maiores produtividades reais nas fazendas Pantanal, Planalto e Paladino, acima de 4500,0 kg ha^{-1} , enquanto as menores produtividades reais foram obtidas nas fazendas Planorte, Parnaíba e Planeste, com médias abaixo de 3900,0 kg ha^{-1} . A fazenda Paladino foi a que apresentou menor precipitação média acumulada nos talhões e os maiores valores de radiação acumulada média, comportamento oposto ao encontrado nos dados da fazenda Paiaguás, que teve os valores de precipitação entre os maiores e os valores de radiação entre os menores.

TABELA 4. Comparativo de médias dos atributos meteorológicos, fitotécnicos e de solo entre as fazendas de produção.

Variáveis	Médias Fazendas ¹								Estatística ANOVA ²			
	PB	PD	PE	PG	PL	PP	PR	PT	F-calc.	CV (%)	p-valor	S-W
<i>Meteorológicos</i>												
Precipitação (mm)	1.025 b	911 c	1.049 b	1.167 a	1.167 a	992 b	992 b	1.062 b	38,31	10,2	2,45e ⁻⁴²	0,0438
Radiação (MJ m ⁻²)	2.376 b	2.562 a	2.355 bc	1.934 d	2.246 c	2.396 b	1.988 d	2.308 bc	126,36	7,2	4,21e ⁻¹⁰³	1,77e ⁻⁰⁸
Época de plantio (dias)	8,45 bc	15,21 ab	-5,13 e	4,15 cd	14,00 ab	16,84 a	-0,38 de	15,94 a	38,8	140	7,42e ⁻⁴³	4,08e ⁻⁹
<i>Fitotécnicos</i>												
Prod. Potencial (kg ha ⁻¹)	6.642 a	6.291 b	6.458 ab	5.733 c	6.307 b	6.496 ab	5.705 c	6.576 ab	52,26	6,84	9,99e ⁻⁵⁵	7,88e ⁻⁷
Prod. Real (kg ha ⁻¹)	3817 de	4505 ab	3623 e	3963 cd	4784 a	4272 bc	3827 de	4823 a	52,40	11,06	7,59e ⁻⁵⁵	2,97e ⁻⁷
YG total (kg ha ⁻¹)	2.825 a	1.855 bc	2.835 a	1.795 bc	1.523 c	2.224 b	1.878 bc	1.754 bc	28,19	32,99	2,57e ⁻³²	2,70e ⁻⁹
YG manejo (kg ha ⁻¹)	2.374 a	1.983 bc	1.364 d	2.141 abc	976 e	2.367 a	2.289 ab	1.859 c	43,06	22,78	1,00e ⁻⁴⁶	1,69e ⁻⁷
<i>Solo</i>												
pH CaCl ₂	4,86 d	5,72 a	4,98 d	5,02 d	5,49 bc	5,39 c	5,62 ab	5,62 ab	109,89	4,68	3,55e ⁻⁹⁴	0,01
H+Al (cmol _c dm ⁻³)	4,54 a	1,69 e	4,60 a	3,99 b	2,49 d	3,10 c	2,32 d	2,13 de	149,29	20,93	2,43e ⁻¹¹⁴	0,002
Al (cmol _c dm ⁻³)	0,05 ab	0,00 c	0,05 ab	0,06 a	0,05 ab	0,00 c	0,01 bc	0,04 abc	9,98	156,54	1,24e ⁻¹¹	6,56e ⁻²¹
Ca (cmol _c dm ⁻³)	3,13 cd	2,18 e	3,41 bc	2,77 d	4,14 a	3,66 ab	3,53 bc	3,88 ab	36,63	22,5	9,73e ⁻⁴¹	1,20e ⁻⁵
Mg (cmol _c dm ⁻³)	0,86 d	0,88 d	0,94 cd	0,89 d	1,25 ab	1,31 a	1,21 ab	1,09 bc	20,48	28,35	5,30e ⁻²⁴	0,001
K (cmol _c dm ⁻³)	0,12 f	0,19 de	0,24 cd	0,16 ef	0,39 a	0,33 ab	0,28 bc	0,35 a	70,66	34,87	4,56e ⁻⁶⁹	6,98e ⁻¹³
P (mg dm ⁻³)	27,98 d	40,71 bc	26,69 d	33,07 cd	42,70 ab	51,34 a	31,89 d	41,11 bc	25,94	34,32	5,87e ⁻³⁰	2,37e ⁻¹²
S (mg dm ⁻³)	8,35 bc	6,30 c	15,14 a	14,35 a	12,79 a	14,34 a	11,16 ab	11,44 ab	16,65	47,71	1,26e ⁻¹⁹	4,10e ⁻¹⁴
B (mg dm ⁻³)	0,87 cd	0,68 d	0,69 d	0,77 cd	1,75 a	0,79 cd	0,97 bc	1,13 b	26,30	36,96	2,48e ⁻³⁰	0,007
Fe (mg dm ⁻³)	48,38 a	27,68 b	42,04 a	30,2 b	30,43 b	29,8 b	26,01 b	25,18 b	36,10	31,69	3,16e ⁻⁴⁰	9,85e ⁻⁰⁸
Mn (mg dm ⁻³)	2,94 c	3,00 bc	3,13 bc	2,90 c	4,58 a	2,95 c	3,93 ab	3,70 abc	6,33	40,93	4,00e ⁻⁷	1,66e ⁻⁷
Cu (mg dm ⁻³)	1,13 b	0,64 c	1,18 ab	1,59 a	1,45 ab	1,51 ab	1,28 ab	1,27 ab	15,16	45,72	7,08e ⁻¹⁸	3,59e ⁻¹⁰
Zn (mg dm ⁻³)	2,49 bc	1,86 c	2,39 bc	2,32 bc	3,66 a	2,59 b	2,55 b	2,85 b	8,46	39,06	9,35e ⁻¹⁰	2,96e ⁻⁹
SB (cmol _c dm ⁻³)	4,11 de	3,24 f	4,58 cd	3,83 ef	5,90 a	5,36 ab	5,06 bc	5,42 ab	38,03	22,52	4,57e ⁻⁴²	1,21e ⁻⁶
CTC (cmol _c dm ⁻³)	8,65 ab	4,91 e	9,17 a	7,82 bcd	8,37 abc	8,45 ab	7,37 d	7,55 cd	60,32	15,41	2,94e ⁻⁶¹	0,001
V (%)	47,54 c	66,23 ab	49,75 c	48,63 c	69,87 a	63,03 b	68,26 ab	71,29 a	111,58	13,05	3,94e ⁻⁹⁵	0,42
MO (dag kg ⁻¹)	2,23 e	1,29 f	3,13 c	2,60 d	3,84 a	2,18 e	3,67 ab	3,35 bc	125,13	18,23	1,84e ⁻¹⁰²	0,007
Argila (dag kg ⁻¹)	33,46 d	18,02 e	59,97 bc	53,71 c	69,89 a	54,69 c	66,65 ab	52,28 c	81,79	24,72	7,56e ⁻⁷⁷	1,69e ⁻¹⁰

Fazendas: Parnaíba (PB), Paladino (PD), Planeste (PE), Paiaguás (PG), Planalto (PL), Pamplona (PP), Planorte (PR), Pantanal (PT). ² F-calc.: valor de F calculado no teste de Fisher; CV: coeficiente de variação; p-valor: valor de p extraído do teste de Fisher à 0,05 de probabilidade; S-W: valor de p extraído do teste de Shapiro-Wilk à 0,05 de probabilidade.

A fazenda de produção Planalto, uma das mais produtivas do conjunto de dados, se destaca pelos parâmetros de solo, apresentando os maiores valores dos teores de Argila e de Matéria Orgânica, além de apresentar valores mais altos das médias de Saturação de bases (V%) e de todos os nutrientes avaliados. A Fazenda Paladino, que figura também entre as de maior produtividade média, tem os menores valores de Argila e de Matéria Orgânica, porém tem altos valores V%, além de apresentar bom teor médio de fósforo (P).

Dentre as fazendas de menores médias de produtividade, a Parnaíba apresenta o menor teor médio de V% entre as fazendas avaliadas, além de apresentar baixos teores de fósforo (P) e de potássio (K), apresentando ainda os maiores valores médios de acidez potencial (H+Al), dos “Yield Gap” total (YGt) e de manejo (YGM).

Na figura 9, são apresentados os valores de coeficientes de correlação de Pearson entre todos os atributos avaliados. Observa-se que a produtividade real (Prod. real) apresentou as maiores correlações positivas com os valores de saturação de bases (V) e de pH do solo, comportamento reforçado pelos resultados da fazenda Planalto, demonstrando que estes atributos vinculados à fertilidade foram relevantes para a variação dos valores de produtividade, reforçando a importância da calagem para o aumento de produtividade (NOLLA & ANGHINONI, 2006). Analisando-se a correlação de produção real com os nutrientes especificamente, percebe-se a importante correlação positiva principalmente do teor de Cálcio (Ca) e Magnésio (Mg), que exercem grande influência no valor de saturação de bases, além de Boro (B), Potássio (K), Fósforo (P), importantes nutrientes para a obtenção de altas produtividades (CASANOVA, 2000; ROSS et al., 2006; SOUSA & LOBATO, 1996).

Quando se avalia a correlação dos valores de produtividade com a época de plantio (época plantio), observa-se uma tendência de maior produtividade da soja em plantios mais

tardios, dentro da realidade dos dados analisados. A menor produção era esperada para as semeaduras mais antecipadas para este conjunto de dados, devido a utilização de cultivares de soja de ciclo mais curto nas primeiras semeaduras do Mato Grosso, em função do cultivo de algodão segunda-safra. Devido ao menor ciclo, a radiação solar acumulada é também menor para estes registros, como observado pelas médias das fazendas Planorte e Paiaguás, impactando no total de fotossíntese e biomassa (KINIRY, 1989; WILSON & JAIESON, 1984). Além disso, no início do período de chuvas, há uma menor regularidade nas precipitações pluviométricas, podendo prejudicar a germinação de sementes e o estande final de plantas.

A alta correlação negativa encontrada entre produtividade real e latitude é possível de ser observada também nos mapas de produtividade dos talhões das fazendas, comparando as produtividades de talhões localizados em maiores latitudes (mais ao sul, com valores mais negativos) como os das fazendas Paiaguás (Apêndice 5 (7-C)), Planalto (Apêndice 4 (6-C)) e Pantanal (Apêndice 4 (5-C)) com talhões localizados em menores latitudes, como os da fazenda Parnaíba (Apêndice 2 (1-C)), por exemplo. A correlação foi negativa também entre a produtividade real e os “Yield Gaps” total e de manejo, demonstrando que as menores produtividades são encontradas em talhões com maior oportunidade de incremento de produção.

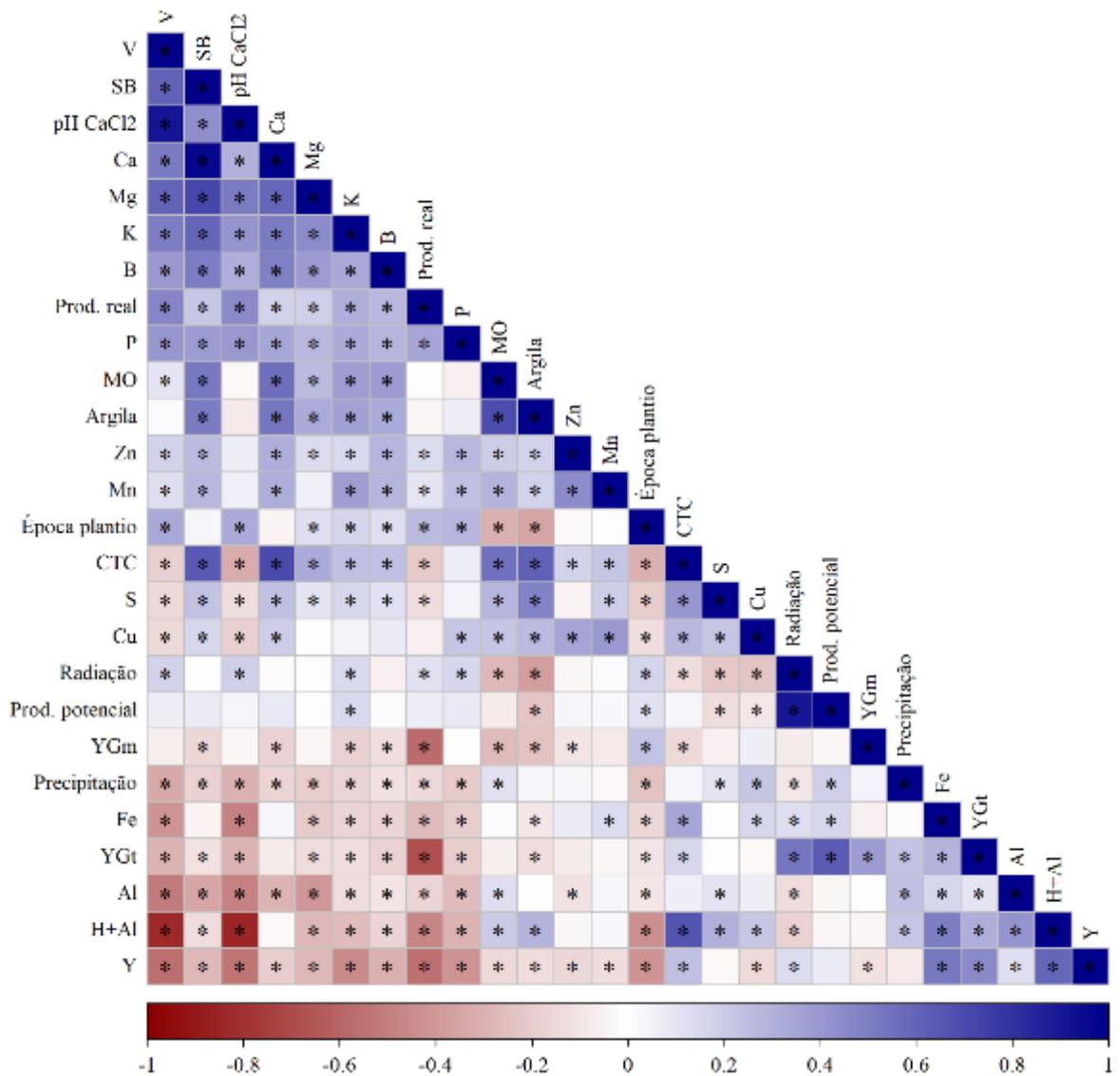


FIGURA 8. Heatmap de coeficientes de correlação de Pearson, obtido a partir de atributos da soja, solo e meteorológicos dos locais de cultivo. * indica correlação significativa ($p\text{-valor} \leq 0,05$). Abreviações: Prod.: Produtividade; YGt: “Yield gap” total; YGm: “Yield gap” de manejo; SB: Soma de bases; V: Saturação de bases; H+Al: Acidez potencial do solo; CTC: Capacidade de troca de cátions; Ca: Cálcio; Mg: Magnésio; K: Potássio; B: Boro; P: Fósforo; MO: Matéria Orgânica; Zn: Zinco; Mn: Manganês; S: Enxofre; Cu: Cobre; Fe: Ferro; Al: Alumínio; Y: Latitude da lavoura.

Apesar da correlação positiva entre produtividade real e época de plantio, o “Yield Gap” de manejo também apresentou correlação positiva com a época, isso pode ser explicado pela grande amplitude geográfica dos dados, de modo que o “Yield Gap” de

manejo local foi aumentado em relação ao atraso para a época ótima, enquanto de modo geral, pelas diferentes amplitudes de produtividades das fazendas de produção, o atraso causou efeito positivo na produtividade. Destacam-se também as correlações negativas do “Yield Gap” de manejo com os valores de argila e de matéria orgânica do solo (MO) e com os valores de soma de bases, pH e potássio, indicando que práticas de manejo relacionadas à fertilidade do solo, como adubação e calagem, tendem a ter impacto na redução do “Yield Gap” e no consequente aumento de produtividade.

As correlações negativas encontradas para precipitação com os atributos de fertilidade do solo (Saturação de bases, pH, potássio, fósforo, entre outros) podem ser explicadas pela realização de plantio de algodão segunda safra nas fazendas do Mato Grosso, que levam ao plantio de cultivares de soja precoce antecipadamente em lavouras de maior potencial produtivo do algodão, de forma que estas lavouras finalizam seu ciclo de soja com uma menor precipitação acumulada, demonstrando que há uma preferência para a produção do algodão em detrimento a altas produtividades de soja. O cultivo de algodão segunda-safra nas lavouras de maior fertilidade ajuda a explicar também a correlação negativa entre a produtividade real da soja e os valores de precipitação acumulada.

4.2 Produtividade Potencial e “Yield Gap”

Os valores de produtividade potencial máximo das fazendas de produção variaram entre 7690,0 kg ha⁻¹ (Fazenda Pamplona - Figura 10-4) e 6557,0 kg ha⁻¹ (Fazenda Planalto - Figura 10-6). Os valores de produtividade atingível variaram entre 6682,0 kg ha⁻¹ (Fazenda Pantanal - Figura 10-5) e 4987,0 kg ha⁻¹ (Fazenda Planeste - Figura 10-2). Quando avaliamos as produtividades reais as médias ponderadas dos talhões variaram entre 5070,0 kg ha⁻¹ (Fazenda Planalto - Figura 10-6) e 3559,0 ka ha⁻¹ (Fazenda Planeste - Figura 10-2).

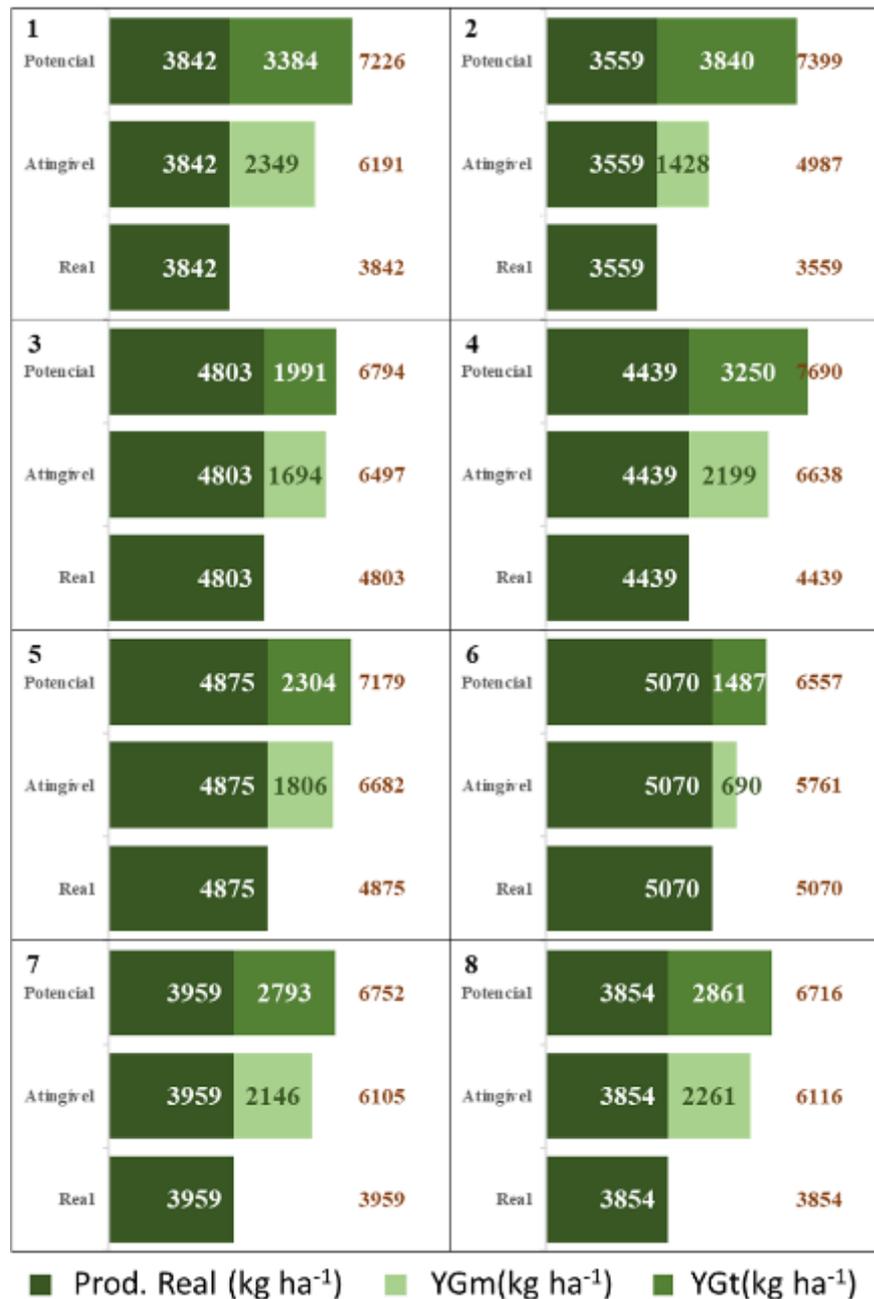


FIGURA 9. Gráficos de produtividades real, atingível e potencial máximo, destacando os valores de “Yield Gap” total (YGt) e “Yield Gap” de manejo (YGm) de cada fazenda de produção. Abreviações: Prod. Real – Produtividade média obtida na safra 2021/22; Identificação: 1: Fazenda Parnaíba; 2 - Fazenda Planeste; 3 – Fazenda Paladino; 4 – Fazenda Pamplona; 5 – Fazenda Pantanal; 6 – Fazenda Planalto; 7 – Fazenda Paiaguás; 8 – Fazenda Planorte.

A fazenda de produção Planalto, devido às características de duração de ciclo das cultivares, dos valores mais baixos de temperatura em relação às demais fazendas e ao valor

de radiação do período avaliado, obteve o menor valor de produtividade potencial máxima, e apresenta a maior produtividade real dentre as fazendas de produção avaliadas, com uma produção de 5070,0 kg ha⁻¹ (Figura 10-6). A mesma fazenda de produção obteve o menor valor de “Yield Gap” total (YGt), sendo de 1487,0 kg ha⁻¹. Para a mesma fazenda de produção, a produtividade atingível foi a segunda menor dentre as avaliadas, com 5761,0 kg ha⁻¹, com o menor valor de “Yield Gap” de manejo (YGm), de 690,0 kg ha⁻¹. Assim, esta fazenda de produção dentre as avaliadas é a que possui a menor margem para o incremento de produtividade vinculado a práticas de manejo, o que pode ser justificado pelo tempo de cultivo e pela estabilidade produtiva, já que a média produtiva está próxima ao máximo de produtividade potencial.

O maior valor de YGm foi encontrado para a Fazenda de produção Parnaíba (Figura 10-1), de 2349,0 kg ha⁻¹, demonstrando o maior espaço para incremento de produtividade vinculado a práticas de manejo das culturas.

A Fazenda de produção Planeste (Figura 10-2), apresentou o maior valor de YGt de 3840,0 kg ha⁻¹, e o YGm foi de 1428,0 kg ha⁻¹, sendo o segundo menor entre as fazendas de produção estudadas. Estes valores demonstram também uma estabilidade de produtividade, porém, com possibilidade de incremento de produtividade com ajustes nas práticas de manejo da cultura da soja.

Os comparativos dos valores de produtividade potencial, produtividade real, YGt e YGm nas Fazendas de produção distribuídas no Brasil podem ser observados na Figura 11. Para o estado do Mato Grosso, há uma semelhança entre os valores das fazendas para todas as variáveis avaliadas, destacando-se entre as menores produtividades potenciais, e os maiores valores de YGm. O estado do Mato Grosso do Sul apresentou os menores valores de produtividade potencial, porém com altos valores de produtividade real, ou seja, obtendo

os menores valores de YGt e de YGm. No estado de Goiás destacam-se os altos valores de produtividade potencial e de YGm, apesar da produtividade real ser alta, porém nesta região podem ser alcançados novos patamares de produtividade com ajustes em práticas de manejo da cultura da soja. No estado do Maranhão encontrou-se as menores produtividades reais, e altos valores de YGt e YGm.

Trabalho realizado por Santos et al. (2021), encontrou valores de YGm entre 2600,0 e 4500,0 kg ha⁻¹ nos estados do Goiás, Bahia e Tocantins. Estes altos valores se devem ao comparativo entre a produtividade real da região com a produtividade atingível baseada em modelagem, cuja única restrição no modelo é a disponibilidade hídrica de cada local. Outro estudo avaliou também o “Yield Gap” de manejo com base em modelagem da produtividade atingível e encontrou valores entre 350,0 e 650,0 kg ha⁻¹ para os estados do Mato Grosso, Mato Grosso do Sul, Goiás, Bahia e Maranhão (SENTELHAS, 2015). Nestes estudos foi direcionado um maior foco para as avaliações de restrição hídrica, não objetivando a identificação dos atributos que impactam o “Yield Gap” de manejo.

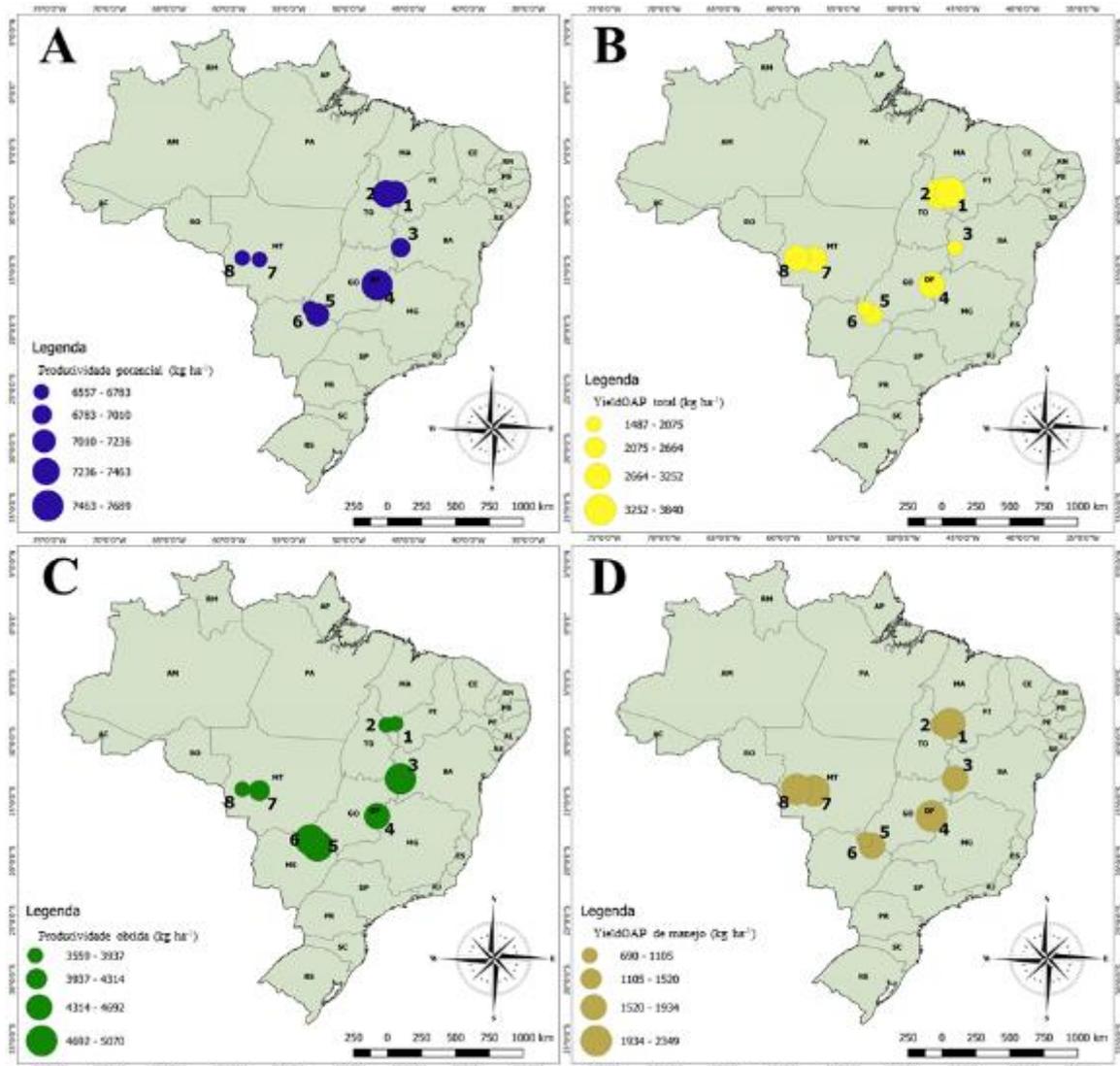


FIGURA 10. Distribuição das médias de (A) produtividade potencial (kg ha^{-1}), (B) “Yield GAP” total (kg ha^{-1}), (C) produtividade de soja obtida na safra 2021/2022 (kg ha^{-1}) e (D) “Yield GAP” de manejo (kg ha^{-1}) no Brasil.

4.3 Análise multivariada

A classificação não supervisionada das lavouras foi realizada pela soma acumulada da distância euclidiana em uma taxa de 67 % do somatório total (Figura 12-B), onde foram separados em quatro grupos para a análise (Figura 12-A).

Na análise de componentes principais (PCA) foram extraídos seis componentes (CP) com autovalores acima de 1, entre 1,4 e 6,4, que juntos explicam 69,8 % da variabilidade

total dos dados coletados e analisados (Tabela 5). Os três primeiros componentes, que juntos explicam mais de 50% da variância, foram selecionados para confecção dos gráficos biplots. O primeiro componente (CP 1) reteve 24,7 % da variância explicada e foi correlacionado positivamente ao atributo H+Al e latitude, e negativamente aos atributos produtividade real, pH, Cálcio (Ca), Magnésio (Mg), Potássio (K), Fósforo (P), Boro (B), soma de bases (SB) e saturação de bases (V), de forma que pode ser entendido como relacionado à fertilidade do solo e produtividade.

O segundo componente (CP 2) reteve 18,2 % da variância explicada e correlacionou-se negativamente com a época de plantio e com o pH. O mesmo componente se relaciona positivamente com o Enxofre (S), CTC, matéria orgânica (MO) e argila, de forma que pode ser relacionado ao posicionamento da cultura da soja nos talhões, que compreende a escolha de semeaduras mais antecipadas em ambientes de esperado melhor potencial produtivo, vinculado aos teores de argila e MO.

O terceiro componente (CP 3) reteve 10,5 % da variância explicada e correlacionou-se negativamente com as variáveis radiação solar, produtividade potencial e com o YGt, reforçando a interação esperada entre estas variáveis, podendo ser relacionado ao potencial produtivo da cultura da soja.

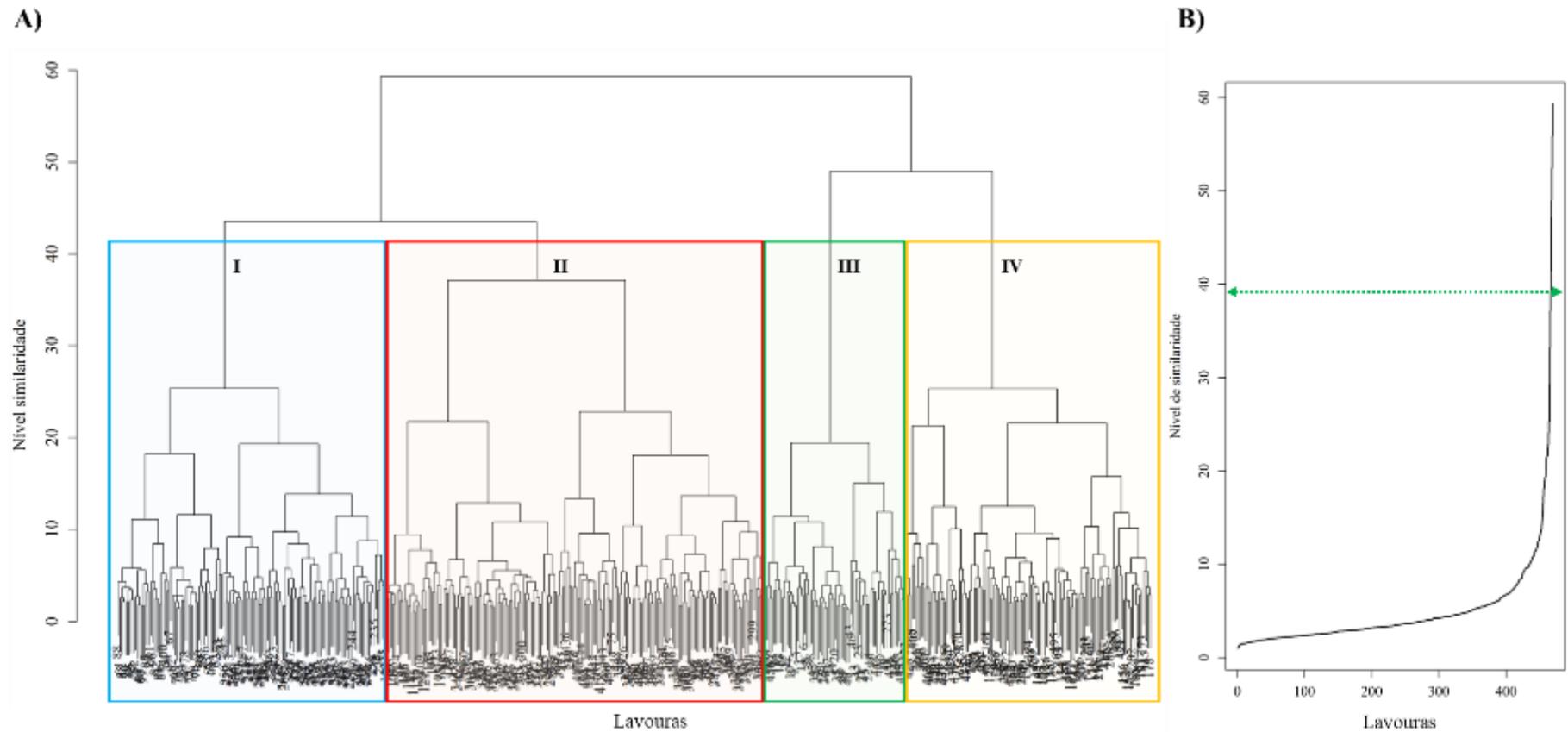


FIGURA 11. Dendrograma com classificação não supervisionada das lavouras demonstrando os grupos (A) e gráfico de distância acumulada demonstrando a altura do corte (B).

TABELA 5. Resumo dos componentes principais (CP) obtidos da análise de componentes principais dos atributos da soja, solo e meteorológicos dos locais de cultivo.

Componente principal	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
Autovalor	6,429	4,720	2,731	1,791	1,679	1,400
Variância explicada (%)	24,7	18,2	10,5	6,9	5,4	4,1
Variância acumulada (%)	24,7	42,9	53,4	60,3	65,7	69,8
Atributo¹	Correlação²					
<i>Meteorológicos</i>						
Precipitação	0,290	0,231	-0,037	0,392	0,467	-0,538*
Radiação	-0,022	-0,368	-0,860*	0,117	0,068	0,127
Época de plantio	-0,266	-0,522*	-0,033	0,175	-0,003	0,381
<i>Fitotécnicos</i>						
Prod. potencial	-0,006	-0,151	-0,893*	0,235	0,238	0,006
Prod. real	-0,539*	-0,362	0,236	0,239	0,176	0,230
YG total	0,389	0,155	-0,801*	0,001	0,037	-0,167
<i>Solo</i>						
pH CaCl ₂	-0,739*	-0,512*	-0,008	-0,143	0,049	-0,169
H+Al	0,546*	0,744	-0,095	-0,012	-0,098	0,227
Al	0,452	0,257	0,136	0,312	0,402	0,376
Ca	-0,754*	0,481	-0,218	-0,156	-0,087	0,007
Mg	-0,711*	0,106	-0,124	-0,345	-0,026	-0,143
K	-0,692*	0,124	-0,153	0,076	0,279	0,215
P	-0,565*	-0,118	-0,076	0,311	-0,291	0,183
S	-0,097	0,546*	0,094	-0,026	0,184	0,140
B	-0,601*	0,192	-0,002	0,085	0,070	0,112
Mn	-0,338	0,305	-0,067	0,555*	-0,218	0,126
Cu	-0,104	0,439	0,124	0,563*	-0,265	-0,215
Zn	-0,340	0,212	-0,031	0,502*	-0,418	-0,237
SB	-0,831*	0,390	-0,212	-0,197	-0,036	-0,007
CTC	-0,206	0,865*	-0,231	-0,156	-0,101	0,170
V	-0,856*	-0,396	-0,060	-0,140	0,022	-0,164
MO	-0,377	0,674*	0,067	-0,024	0,319	-0,085
Argila	-0,365	0,739*	0,171	-0,138	0,175	-0,041
<i>Localização</i>						
Latitude	0,625*	0,259	-0,320	-0,353	-0,370	0,081

¹Abreviações: Prod.: Produtividade; YGt: “Yield gap” total; YGm: “Yield gap” de manejo; SB: Soma de bases; V: Saturação de bases; H+Al: Acidez potencial do solo; CTC: Capacidade de troca de cátions; Y: Latitude da lavoura.

²Correlação: * atributo significativo no componente.

Observa-se pela variância explicada pelos atributos no biplot dos componentes 1 e 2 (Figura 13-A), relacionados a fertilidade do solo, produtividade real e ao posicionamento da cultura nos talhões, que a época de plantio tem valores opostos aos teores de MO e argila, evidenciando que as semeaduras mais tardias (final do zoneamento) acontecem em lavouras com menores teores destes atributos (MO e argila), haja visto que o algodão segunda-safra é posicionado nas lavouras com maiores valores destes atributos, forçando a utilização de cultivares de soja de ciclo mais curto e nas épocas mais iniciais. Fica evidenciado que os valores de produtividade real e das variáveis P e V% têm alta relação, corroborando com o incremento de produtividade em virtude da fertilidade do solo.

O teor de argila do solo aparece em sentido oposto à produtividade real, o que pode ser exemplificado pelos dados da Fazenda de produção Paladino (PD), com baixo teor de argila e altos valores de produtividade real. A Fazenda de produção Pantanal (PT) obteve correlação com os valores de P, V% e produtividade real. Para a Fazenda de produção Planeste (PE) observa-se uma proximidade com Argila, MO e CTC, e grande afastamento da produtividade real e das variáveis P e V%.

Analisando as Fazendas de produção Planorte (PR) e Paiaguás (PG), no Mato Grosso, observou-se um grande afastamento das cargas do autovetor radiação solar acumulada, explicado pela realização do cultivo de algodão segunda safra, sendo que essas áreas precisam ser semeadas com cultivares de soja de ciclo precoce. (Figura 13-B). Porém, devido à menor quantidade de radiação solar acumulada nestas áreas, por consequência a produtividade potencial e o YGt aparecem também com cargas de autovetores opostos.

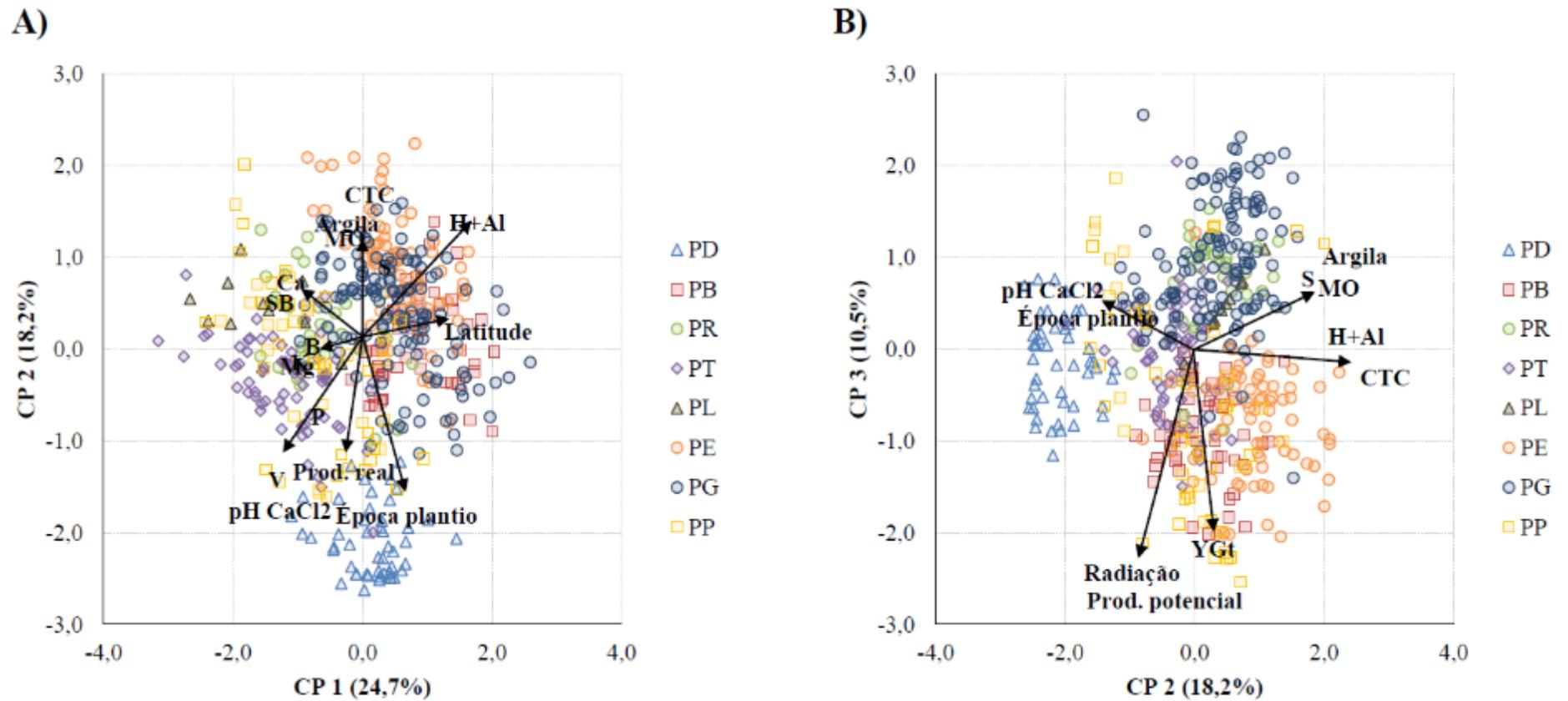


FIGURA 12. PCA biplots dos componentes principais 1 e 2 (A) e componentes principais 2 e 3 (B), indicando as áreas de estudo (pontos) e os carregamentos de cada atributo (setas). Os pontos próximos correspondem a observações com pontuações semelhantes nos componentes da PCA. O comprimento das setas indica a variação dos atributos no componente, enquanto os pontos entre eles indicam suas correlações. Abreviações: PD – Paladino; PB – Parnaíba; PR – Planorte; PT – Pantanal; PL – Planalto; PE – Planeste; PG – Paiaguás; PP – Pamplona.

A análise dos grupos hierárquicos nos biplots do PCA permitiu identificar os principais atributos que influenciam no agrupamento das lavouras (Figura 14). O grupo I é composto principalmente por talhões das fazendas de produção Parnaíba e Planeste (Tabela 6), e tem como atributos de influência CTC, teor de Argila, MO e acidez potencial, além do distanciamento do atributo produtividade real, demonstrando ser um grupo composto por talhões com menores produtividades. Estes mesmos atributos influenciaram no grupo II, composto principalmente por talhões das fazendas de produção Paiaguás e Planorte. Este grupo aparece sobreposto ao grupo I no biplot dos PC1 e PC2 (Figura 14-A). A diferença entre os grupos I e II é possível de ser observada no biplot dos PC 2 e PC3 (Figura 14-B), que demonstra para o grupo I maior radiação, maior produtividade potencial e maior YGt. Portanto, conclui-se que ambos os grupos I e II têm espaço para melhoria de fertilidade e aumento de produtividade real, embora o aumento de produtividade para o grupo II seja também restrito à radiação, que é menor devido ao emprego de cultivares de ciclo mais curto em função do cultivo de algodão segunda-safra no estado do Mato Grosso.

O Grupo III é composto pela fazenda de produção Paladino e por 20% dos talhões da fazenda de produção Pamplona, e tem como atributos de influência principais a V%, produtividade real, pH, P e época de plantio, evidenciando também o afastamento dos atributos Argila e MO. Este grupo representa os talhões arenosos com alta produtividade de soja, destacando a influência positiva dos atributos de fertilidade do solo. Vale destacar que nesta região não ocorre o plantio de segunda-safra, o que explica a influência do atributo época de plantio.

O Grupo IV é composto pelos talhões das fazendas de produção Planalto, Pantanal e Pamplona, e têm como influência a menor latitude, os teores de Argila e de MO, e principalmente os atributos de fertilidade do solo, como Cálcio, Fósforo, Magnésio, Boro e

V%. Este grupo, portanto, separou os talhões de alta produtividade vinculadas a maiores teores de argila e boa fertilidade do solo. É separado do grupo III principalmente devido aos teores de argila e ao atributo época de plantio.

Na análise realizada pelo método dendrograma houve comportamento diferenciado para os talhões das fazendas de produção avaliadas, sendo que 35,9% dos talhões avaliados se enquadraram no grupo II, composto majoritariamente pelos talhões da fazenda de produção Paiaguás (Tabela 6).

TABELA 6. Contagem de talhões das fazendas em cada grupo do dendrograma.

Fazenda	Grupo I	Grupo II	Grupo III	Grupo IV	Talhões / fazenda
Parnaíba	50	2	0	0	52
Planeste	72	3	1	0	76
Paladino	0	0	48	0	48
Pamplona	0	9	10	33	52
Pantanal	0	1	4	63	68
Planalto	0	0	0	10	10
Paiaguás	0	125	3	2	130
Planorte	1	29	0	4	34
Total de talhões	123	169	66	112	470
Porcentagem talhão/grupo (%)	26,1	35,9	14,0	23,8	100

O comparativo de médias dos grupos hierárquicos pode ser observado na tabela 7. O grupo I apresenta a menor média de produtividade, reforçando seu distanciamento deste vetor no gráfico biplot. Observam-se ainda os menores valores de V%, de P e de K, e o maior valor de H+Al para este grupo. O Grupo II apresenta a menor média de radiação acumulada dos talhões, o que vai ao encontro da separação observada entre os grupos I e II no gráfico representando os PC 2 e PC 3. Além disso o grupo II apresenta maior teor de argila, de V%, de P e menor teor de K em relação ao grupo I.

O grupo III separou talhões que receberam as maiores médias de radiações e obtiveram altas médias de produtividade real. Neste grupo encontram-se talhões com altos valores de V%, e os menores valores de H+Al. Destaca-se ainda, para o grupo III, o alto valor de “Yield Gap” de manejo.

O Grupo IV separou os talhões de altas médias de produtividade real, e com os melhores valores médios dos atributos vinculados à fertilidade do solo, como pH e V%, altos teores de Argila e MO além dos maiores teores da maioria dos macro e micronutrientes do solo, justificando o posicionamento deste grupo no gráfico biplot dos componentes principais 1 e 2.

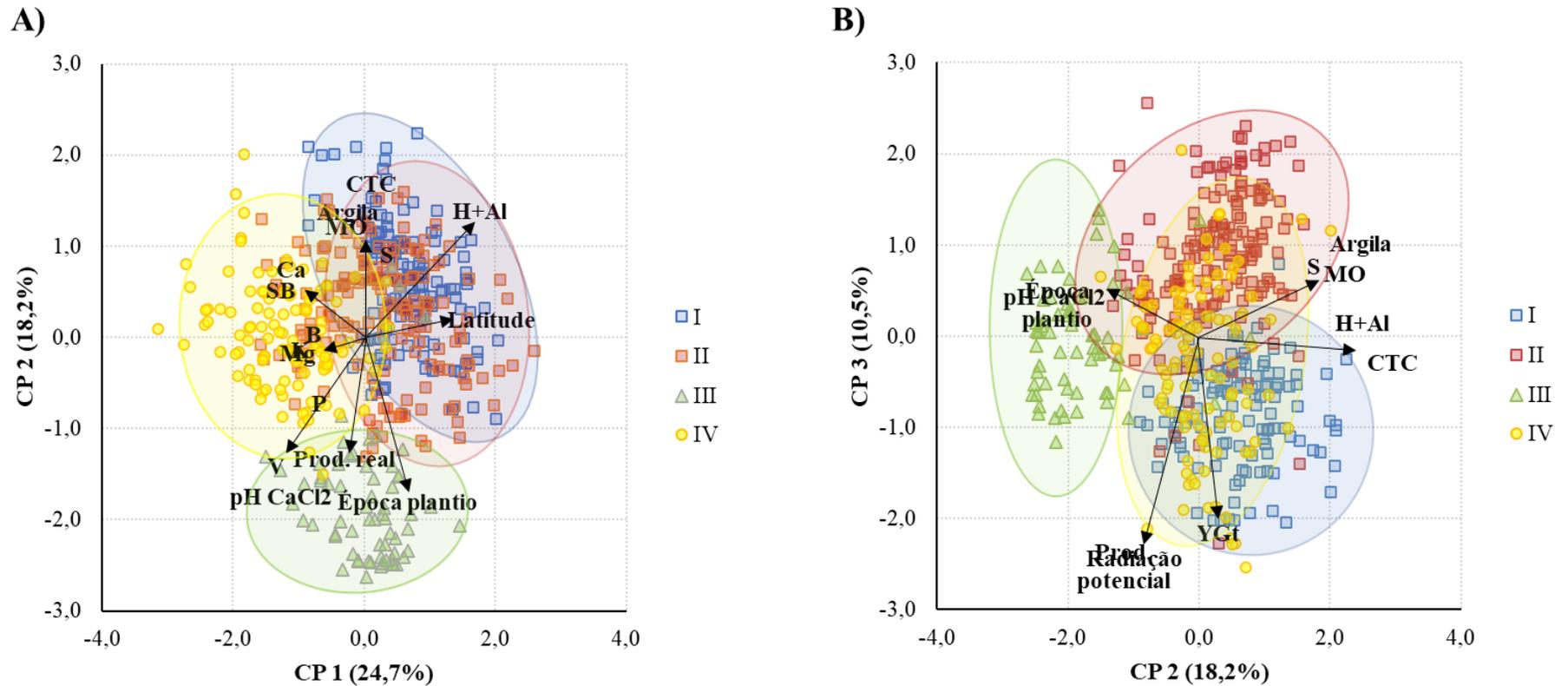


FIGURA 13. PCA biplots dos componentes principais 1 e 2 (A) e componentes principais 2 e 3 (B), indicando a classificação de grupos da análise hierárquica (pontos) e os carregamentos de cada atributo (setas). Os pontos próximos correspondem a observações com pontuações semelhantes nos componentes da PCA. O comprimento das setas indica a variação dos atributos no componente, enquanto os pontos entre eles indicam suas correlações.

TABELA 7. Comparativo de médias e análise ANOVA dos atributos meteorológicos, fitotécnicos e de solo entre os grupos separados pela classificação não supervisionada.

Variáveis	Médias grupos				Estatística ANOVA ¹			
	Grupo I	Grupo II	Grupo III	Grupo IV	F-calc.	CV (%)	p-valor	S-W
<i>Meteorológicos</i>								
Precipitação (mm)	1042 b	1118 a	926 c	1066 b	41,20	11,38	1,23e ⁻²³	5,34e ⁻⁵
Radiação (MJ m ⁻²)	2367 b	1963 c	2464 a	2348 b	200,22	8,08	2,07e ⁻⁸³	6,45e ⁻⁸
Época de plantio (dias)	-0,14 c	3,14 c	18,77 a	14,76 b	73,88	145,64	4,16e ⁻³⁹	5,69e ⁻⁷
<i>Fitotécnicos</i>								
Prod. potencial (kg ha ⁻¹)	6545 a	5750 c	6253 b	6571 a	115,95	6,9	4,28e ⁻⁵⁶	5,78e ⁻⁶
Prod. real (kg ha ⁻¹)	3717 c	3950 b	4451 a	4598 a	78,64	12,02	3,55e ⁻⁴¹	3,60e ⁻⁷
YG total (kg ha ⁻¹)	2828 a	1820 b	1853 b	1973 b	54,4	33,7	3,65e ⁻³⁰	1,59e ⁻¹⁰
YG manejo (kg ha ⁻¹)	1769 c	2170a	2032ab	1958 bc	12,44	28,05	7,68e ⁻⁸	0,0001
<i>Solo</i>								
pH CaCl ₂	4,93 c	5,15 b	5,61 a	5,53 a	116,13	5,75	3,67e ⁻⁵⁶	4,29e ⁻⁵
H+Al (cmol _c dm ⁻³)	4,61 a	3,60 b	1,94 d	2,61 c	168,72	26,06	4,86e ⁻⁷⁴	0,001
Al (cmol _c dm ⁻³)	0,05 a	0,05 a	0,01 b	0,03 b	10,57	161,82	9,72e ⁻⁷	3,41e ⁻²⁴
Ca (cmol _c dm ⁻³)	3,30 b	2,94 c	2,41 d	3,94 a	68,94	23,25	6,51e ⁻³⁷	0,0002
Mg (cmol _c dm ⁻³)	0,90 b	0,97 b	0,91 b	1,20 a	22,95	30,16	7,08e ⁻¹⁴	3,01e ⁻⁵
K (cmol _c dm ⁻³)	0,19 bc	0,18 c	0,22 b	0,36 a	100,05	38,97	5,30e ⁻⁵⁰	4,10e ⁻¹³
P (mg dm ⁻³)	27,13 d	32,25 c	41,16 b	46,52 a	58,13	34,41	6,12e ⁻³²	4,20e ⁻¹²
S (mg dm ⁻³)	11,90 a	13,81 a	6,80 b	13,73 a	23,64	49,53	2,93e ⁻¹⁴	1,44e ⁻¹⁶
B (mg dm ⁻³)	0,77 bc	0,80 b	0,67 c	1,12 a	33,49	39,48	1,28e ⁻¹⁹	0,0003
Fe (mg dm ⁻³)	45,09 a	28,99 b	27,4 b	28,24 b	75,62	32,18	7,17e ⁻⁴⁰	4,14e ⁻¹⁰
Mn (mg dm ⁻³)	3,10 b	2,88 b	2,72 b	4,02 a	22,45	39,88	1,34e ⁻¹³	8,12e ⁻⁷
Cu (mg dm ⁻³)	1,18 b	1,45 a	0,71 c	1,53 a	33,20	45,82	1,82e ⁻¹⁹	2,02e ⁻¹⁰
Zn (mg dm ⁻³)	2,46 b	2,30 b	1,93 c	3,00 a	24,24	35,47	1,35e ⁻¹⁴	3,56e ⁻⁷
SB (cmol _c dm ⁻³)	4,39 b	4,10 b	3,53 c	5,60 a	68,17	23,47	1,44e ⁻³⁶	9,64e ⁻⁵
CTC (cmol _c dm ⁻³)	8,99 a	7,69 c	5,47 d	8,20 b	116,25	16,06	3,28e ⁻⁵⁶	0,001
V (%)	48,65 c	53,21 b	65,15 a	68,06 a	112,03	16,25	1,25e ⁻⁵⁴	0,68
MO (dag kg ⁻¹)	2,81 b	2,71 b	1,55 c	3,16 a	84,23	24,88	1,48e ⁻⁴³	2,07e ⁻⁶
Argila (dag kg ⁻¹)	49,65 b	55,25 a	24,34 c	57,66 a	82,12	29,49	1,15e ⁻⁴²	0,003

¹ F-calc.: valor de F calculado no teste de Fisher; CV: coeficiente de variação; p-valor: valor de p extraído do teste de Fisher à 0,05 de probabilidade; S-W: valor de p extraído do teste de Shapiro-Wilk à 0,05 de probabilidade

4.4 Modelagem da produtividade da soja

O modelo Random Forest (RF) para a estimativa da produtividade da cultura da soja foi ajustado aos dados de treinamento com ótima precisão ($R = 0,97$) (Figura 15). As 352 amostras (75% do conjunto de dados) usadas para treinamento, foram suficientes para ajustar o modelo nas condições do estudo. Quando se testou o restante do conjunto de dados (118 amostras) encontrou-se correlação moderada com $R = 0,59$. Além disso, observa-se na figura 14 a tendência de acompanhamento da linha 1:1 pelo conjunto de teste. Observa-se ainda que os maiores afastamentos dos dados de teste da linha 1:1 ocorrem em zonas de produtividades obtidas com menor quantidade de dados de treinamento, principalmente abaixo de 3000 kg ha^{-1} . As métricas de avaliação do modelo podem ser observadas na tabela 8. A partir da modelagem da produtividade da soja, foi possível realizar a análise da importância relativa das variáveis.

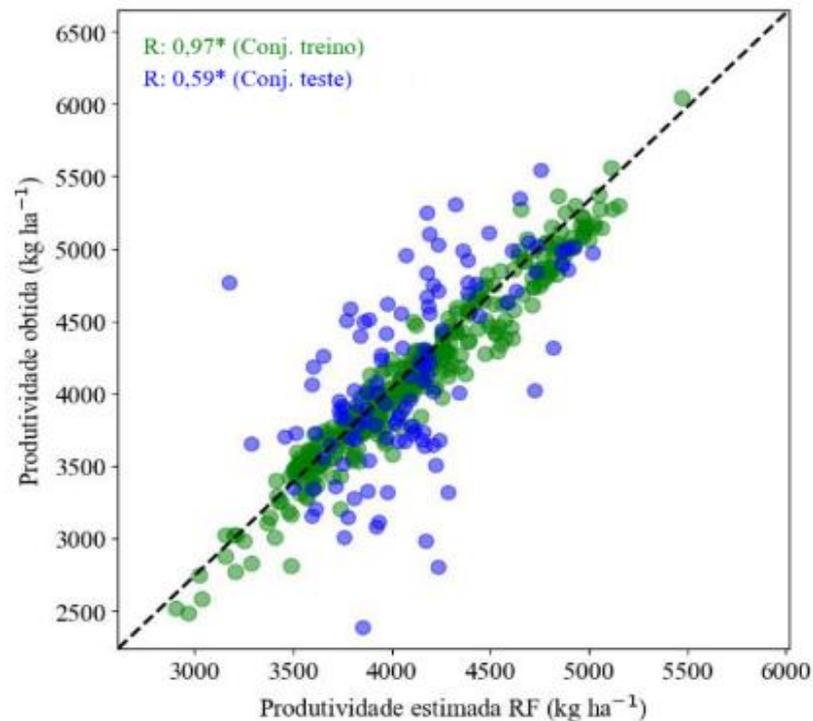


FIGURA 14. Produtividade real estimada pelo modelo Random Forest (RF) em função da produtividade real obtida (observada). Conjunto de dados de treinamento (em verde): 352 observações e; conjunto de dados de teste (em azul): 118 observações.

TABELA 8. Métricas de avaliação do modelo de *machine learning* Random Forest.

Modelo ¹	R ²	MAE	RMSE	N° variáveis de entrada	N° de árvores aleatórias
Random Forest	0,94	165,001	246,043	21	1.000

¹ MAE: erro absoluto médio (mean absolute error); RMSE: raiz quadrada do erro médio (root mean square error).

4.5 Importância relativa das variáveis

A análise da importância relativa das variáveis (IRV) demonstra que o grupo de fertilidade do solo é responsável por 40 % da importância na modelagem da produtividade da soja, seguido pelos grupos de clima, macronutrientes e micronutrientes, com 22 %, 17 % e 17 %, respectivamente (Figura 16-A).

Quando se analisa o grupo de fertilidade do solo, observa-se que os atributos saturação de bases (V%) e H+Al apresentam as maiores contribuições dentro do grupo, explicando 18 % e 13 %, respectivamente das variações (Figura 16-B). Os demais atributos como pH CaCl₂, MO, CTC, Al e SB contribuíram com valores inferiores a 3 %.

No grupo de clima (Figura 16-C), a variável radiação teve a maior importância relativa, com 12 %, seguido por precipitação (6 %) e época de plantio (4 %). A época de plantio foi a principal causa de perda de produtividade vinculada a manejo encontrada por Tagliapietra et al. (2021), em estudo realizado no sul do Brasil. Este atributo não teve maior relevância para o conjunto de dados pois já é de prática da SLC Agrícola realizar as semeaduras nas melhores janelas para a cultura, de forma que é um atributo do “Yield Gap” de manejo já mitigado.

No grupo de macronutrientes, a maior importância relativa foi encontrada para a variável P com 8 %, seguido por S com 4 %, Mg e K ambos com 2 %, e Ca com 1 % (Figura 16-D). As variáveis de maior importância do grupo de micronutrientes foram Fe Mn com 4 % cada um, seguidos por B, Cu e Zn ambos com 3 % (Figura 16-E).

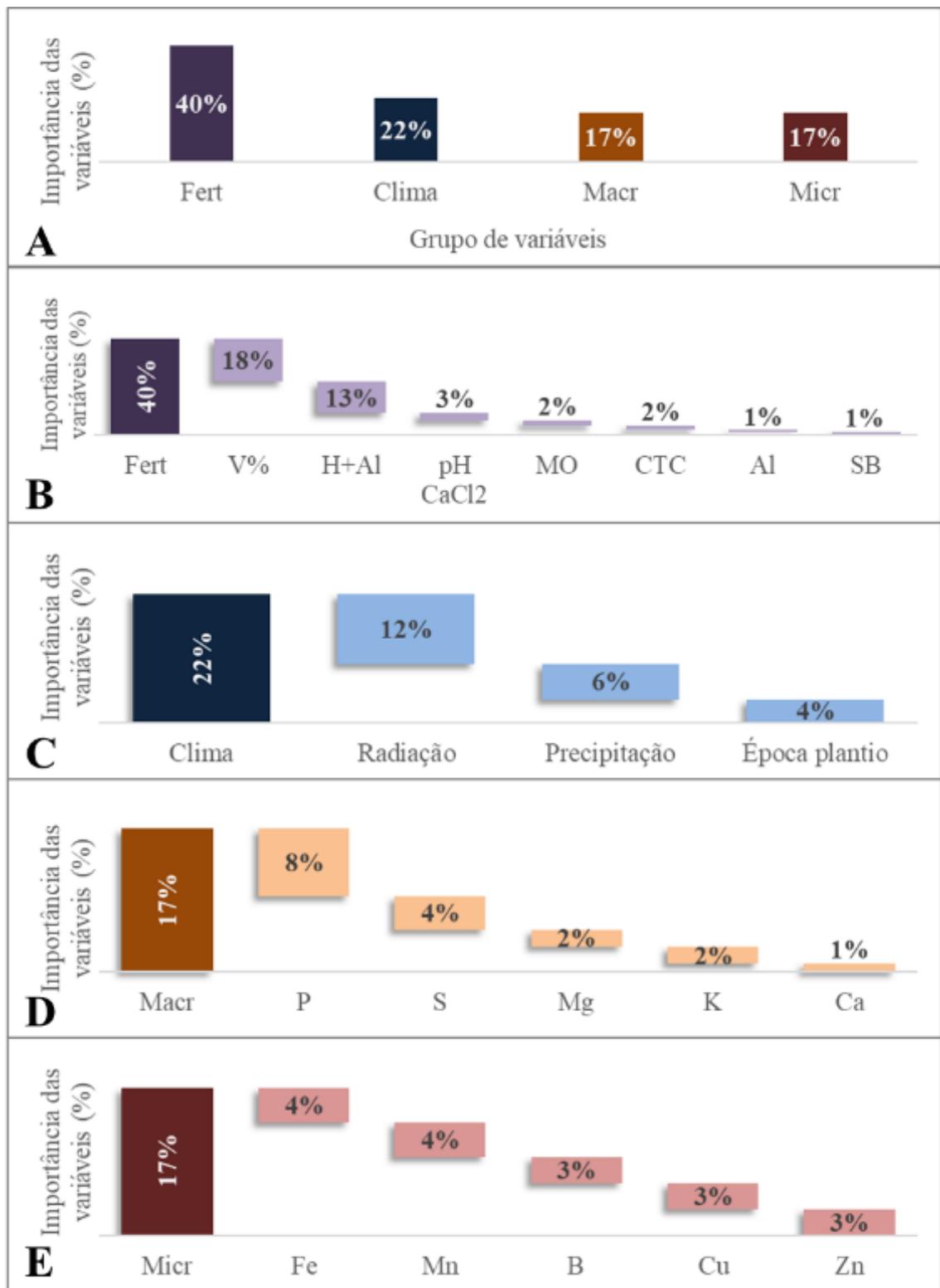


FIGURA 15. Importância relativa das variáveis avaliadas na redução de produtividade da Soja, agrupados por (A) Grupo de variáveis e aberto nos grupos de (B) Fertilidade do solo, (C) Climáticas, (D) Macronutrientes e (E) Micronutrientes. Abreviações: Fert – Fertilidade do solo; Macr – Macronutrientes do solo; Micr – Micronutrientes do solo.

Ordenando-se todas as variáveis pelo valor de importância relativa, observa-se que as variáveis que se destacaram com maiores importâncias foram V%, seguido por H+Al, ambos atributos de fertilidade do solo, seguidas pela radiação solar, pelo P e pela precipitação (Figura 17). Juntas, estas cinco variáveis retêm 57% da importância relativa das variáveis da modelagem de produtividade de soja.

Os atributos de fertilidade podem ser controlados pela correção e adubação, como aplicação de fertilizantes fosfatados e a aplicação de calcário. Os teores de P, a saturação de bases e a acidez potencial apresentaram forte correlação de Person com a produtividade real do conjunto de dados, e tiveram ainda sua importância reforçada pelos seus autovetores do CP 1, demonstrando serem fatores determinantes para a produtividade da soja. Já a precipitação e a radiação têm ligação com o ciclo das cultivares, cuja escolha também é considerada prática de manejo da cultura. O aumento de fertilidade das áreas de produção é citado também por Mueller et al. (2012), como prática capaz de reduzir os gaps de produtividade e aumentar a eficiência produtiva das culturas nos próximos anos. Os autores citam também, a irrigação como alternativa para reduzir o gap produtivo em função da precipitação.

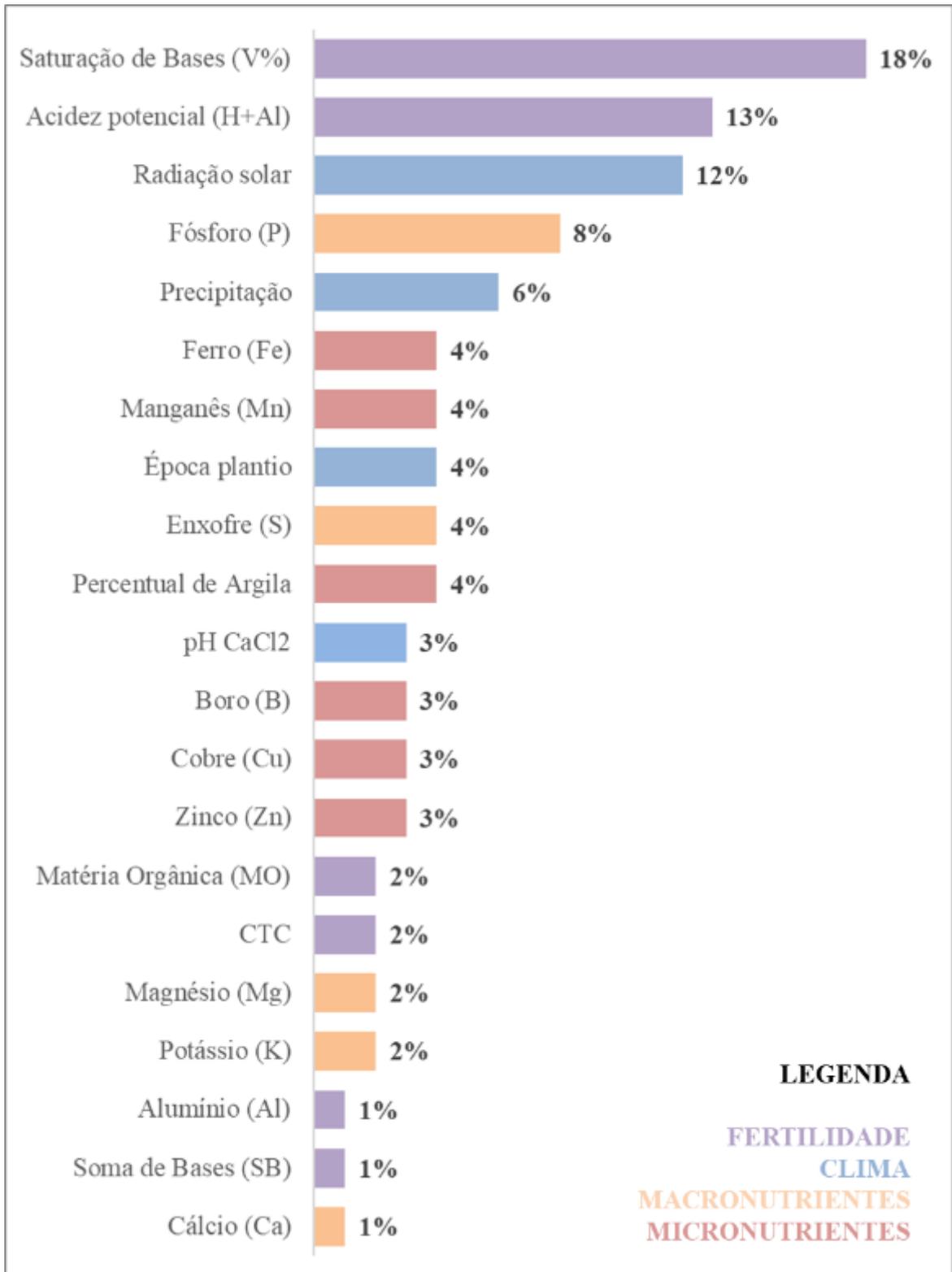


FIGURA 16. Importância relativa das variáveis avaliadas no modelo de produtividade da Soja, classificadas pela ordem de maior para o menor impacto na produtividade.

5. CONCLUSÃO

A produtividade de soja foi fortemente impactada pelos atributos associados ao solo, e as análises demonstram que a saturação de bases representou 18 % da importância dentre as variáveis observadas, seguida pela acidez potencial, com 13 % e os teores de fósforo, com 8 %. Entre os atributos de clima, destacam-se a radiação solar, com 12 % e a precipitação, com 6 %. Estas variáveis podem ser manejadas com práticas de calagem e adubação, e principalmente pelo ajuste do ciclo das culturas e época de plantio, possibilitando redução do “Yield Gap” de manejo e aumento de produtividade.

A distribuição geográfica das áreas de estudo e a variabilidade dos ambientes de produção proporcionou ajustar um modelo de produtividade e permitiu uma robusta avaliação dos atributos que impactam na produtividade da cultura da soja, desvinculando o resultado dos efeitos locais de cada fazenda. O fluxo de análise dos dados proposto neste estudo foi capaz de identificar de forma coerente os principais atributos associados à variação de produtividade, configurando alternativa viável para o entendimento do “Yield Gap” de manejo para um conjunto de dados amplo e com variabilidade espacial.

As análises permitiram ainda inferir sobre as características semelhantes entre os talhões mais produtivos e as semelhanças também entre os talhões menos produtivos, possibilitando a análise dentro dos grupos hierárquicos por similaridade.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. de M.; SPAROVEK, G. Koppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, Stuttgart, v. 22, p. 711–728, 2013.

AGÊNCIA NACIONAL DE ÁGUAS - ANA. **Precipitação Média Anual - Série de 1977 a 2006**. Brasília, DF: ANA, 2016. Disponível em: <https://dadosabertos.ana.gov.br/search>. Acesso em 01 out. 2021.

BANDARA, A. Y.; WEERASOORIYA, D. K.; MURILLO-WILLIAMS, A.; WHITE, C. M.; COLLINS, A. A.; BELL, T. H.; ESKER, P. D. Relationship between soybean yield from high and low yielding field sites and selected soil characteristics. **Agrosystems, Geosciences & Environment**, Hoboken, v. 3, [art.] e20126, 2020.

BATTISTI, R.; SENTELHAS, P. C.; PASCOALINO, J. A. L.; SAKO, H.; DANTAS, J. P. de S.; MORAES, M. F. Soybean Yield gap in the áreas of yield contest in Brazil. **International Journal of Plant Production**, Cham, v. 12, p. 159-168, 2018.

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

CASANOVA, E. Phosphorus and potassium fertilization and mineral nutrition of soybean. **Interciência**, Caracas, v. 25, n. 2, p. 92-95, 2000.

CAMBARDELLA, C.A., MOORMAN, T. B., NOVAK, J. M. PARKIN, T. B. KARLEN, F. L., TURCO, R. F., KONOPKA, A. E. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, Madison, v. 58, n. 5, p. 1501-1511, 1994.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA - CEPEA. **Mercado de trabalho do agronegócio**. São Paulo: CEPEA, [2020]. Disponível em: <https://www.cepea.esalq.usp.br/br/mercado-de-trabalho-do-agronegocio.aspx>. Acesso em: 30 set. 2020.

COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Série histórica das safras**. Brasília, DF: CONAB, [2022]. Disponível em: <https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras>. Acesso em: 12 out. 2022.

CORTEZ, J. W.; NAGAHAMA, H. D. J.; OLSZEWSKI, N.; PATROCINIO FILHO, A. P.; SOUZA, E. B. Umidade e temperatura de Argissolo amarelo em sistemas de preparo e estádios de desenvolvimento do milho. **Journal of the Brazilian Association of Agricultural Engineering**, Jaboticabal, v. 35, p. 699-710, 2015.

CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, Maringá, v. 35, n. 3, p. 271-276, 2013.

REIS, J. C. dos; RODRIGUES, G. S.; DE BARROS, I.; RODRIGUES, R. D. A. R.; GARRETT, R. D.; VALENTIM, J. F.; KAMOI, M. Y. T.; MICHETTI, M.; WRUCK, F. J.; RODRIGUES-FILHO, S.; PIMENTEL, P. E. O.; SMUKLER, S. Integrated crop-livestock systems: a sustainable land-use alternative for food production in the Brazilian Cerrado and Amazon. **Journal of Cleaner Production**, Amsterdam, v. 283, 2021.

SANTOS, H. G. dos; JACOMINE, P. K. T.; ANJOS, L. H. C. dos; OLIVEIRA, V. A. de; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A. de; ARAUJO FILHO, J. C. de; OLIVEIRA, J. B. de; CUNHA, T. J. F. **Sistema Brasileiro de Classificação de Solos**. 5. ed. Brasília, DF: Embrapa, 2018.

FAO and DWFL. Yield gap analysis of field crops – Methods and case studies. **FAO Water reports**, v. 41, p. 1–82, 2015.

GUILPART, N.; GRASSINI, P.; SADRAS, V.O.; TIMSINA, J.; CASSMAN, K.G. Estimating yield gaps at the cropping system level. **Field Crops Research**, Amsterdam, v. 206, p. 21-32, 2017.

FIGUEIREDO, C. C. D.; SANTOS, G. G.; PEREIRA, S.; NASCIMENTO, J. L. D.; ALVES JÚNIOR, J. Propriedades físico-hídricas em Latossolo do Cerrado sob diferentes sistemas de manejo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, p. 146-151, 2009.

HARGREAVES, G. H.; SAMANI, Z. A. Estimating potential evapotranspiration. **Journal of Irrigation and Drainage Engineering**, Reston, VA, v. 108, p. 225-230, 1982.

HOURLBORG, R; McCABE, M.F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 135, p. 173-188, 2018.

IATROU, M.; PAPADOPOULOS, A.; PAPADOPOULOS, F.; DICHALA, O.; PSOMA, P.; BOUNTLA, A. Determination of Soil-Available Micronutrients using the DTPA and Mehlich 3 Methods for Greek Soils Having Variable Amounts of Calcium Carbonate. **Communications in Soil Science and Plant Analysis**, New York, v. 46, p. 1905-1912, 2015.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Biomass e Sistema Costeiro-Marinho do Brasil - 1:250000**. Brasília, DF: IBGE, 2019. Disponível em: <https://www.ibge.gov.br/pt/inicio.html>. Acesso em: 01 out. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Censo Agropecuário de 2017**. Brasília, DF: IBGE, 2017. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria>. Acesso em: 01 out. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Mapa político do Brasil 2020**. Brasília, DF: IBGE, 2020. Disponível em: <https://portaldemapas.ibge.gov.br/portal.php#mapa222624>. Acesso em: 20 out. 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE². **Produção agrícola municipal**. Brasília, DF: IBGE, [2020]. Disponível em: <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>. Acesso em: 28 set. 2020.

ITTEKSUM, M. K. van; CASSMAN, K. G. Yield gap analysis: rationale, methods and applications – Introduction to the special issue. **Field Crops Research**, Amsterdam, v. 143, p. 1-3, 2013.

ITTEKSUM, M. K. van; CASSMAN, K. G.; GRASSINI, P.; WOLF, J.; TITTONELL, P.; HOCHMAN, Z. Yield gap analysis from local to global relevance – a review. **Field Crops Research**, Amsterdam, v. 143, p. 4-17, 2013.

KINIRY, J. R.; JONES, C. A.; O'TOOLE, J. C.; BLANCHET, R.; CABELGUENNE, M.; SPANIEL, D. A. Radiation-use efficiency in biomass accumulation prior to grain filling in five grain-crop species. **Field Crops Research**, Amsterdam, v. 20, p. 51-64, 1989.

MENDES, I. de C.; SOUZA, L. M. de; SOUSA, D. M. G. de; LOPES, A. A. de C.; JUNIOR, F. B. de R.; LACERDA, M. P. C.; MALAQUIAS, J. V. Critical limits for microbial indicators in tropical Oxisols at post-harvest: The FERTBIO soil sampling concept. **Applied Soil Ecology**, Amsterdam, v. 139, p. 85-93, 2019.

MUELLER, N. D.; GERBER, J. S.; JOHNSTON, M.; RAY, D. K.; RAMANKUTTY, N.; FOLEY, J. A. Closing yield gaps through nutrient and water management. **Nature**, London, v. 490, p. 254–257, 2012.

MURPHY, J.; RILEY, J. P. A modified single solution method for the determination of phosphate in natural waters. **Analytica Chimica Acta**, Amsterdam, v. 27, p. 31-36, 1962.

NÓIA JÚNIOR, R. de S.; SENTELHAS, P. C. Yield gap of double-crop system of main-season soybean with off-season maize in Brazil. **Crop & pasture Science**, Collingwood, v. 71, p. 445-458, 2020.

NOLLA, A.; ANGHINONI, I. Critérios de calagem para a soja no sistema plantio direto consolidado. **Revista brasileira de ciência do solo**, Viçosa, MG, v. 30, p. 475-483, 2006.

OLIVEIRA, G. C.; JUNIOR, D.; RESCK, D. V. S.; CURI, N. Caracterização química e físico-hídrica de um Latossolo Vermelho após vinte anos de manejo e cultivo do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, MG, v. 28, p. 327-336, 2004.

OLIVEIRA, G. D. L. The geopolitics of Brazilian soybeans. **The Journal of Peasant Studies**, London, v. 43, p. 348-372, 2016.

OLIVEIRA, G. H.; BAYMA-SILVA, G. Land-use and land-cover mapping of the brazilian cerrado based mainly on landsat-8 satellite images. **Revista Brasileira de Cartografia**, Rio de Janeiro, v. 69, p. 1041-1051, 2017.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Áustria, 2021. Disponível em: <https://www.R-project.org/>. Acesso em: 17 set. 2021.

ROSS, J. R.; SLATON, N. A.; BRYE, K. R.; DELONG, R. E. Boron fertilization influences on soybean yield and leaf and seed boron concentrations. **Agronomy Journal**, Madison, v. 98, p. 198-205, 2006.

SANTOS, T. G.; BATTISTI, R.; CASAROLI, D.; ALVARES JÚNIOR, J.; EVANGELISTA, A. W. P. Assessment of agricultural efficiency and yield gap for soybean um the brazilian central Cerrado biome. **Bragantia**, Campinas, v. 80, [art.] e1821, 2021.

SCARAMUZZA, C. A. M.; SANO, E. E.; ADAMI, M.; BOLFE, E. L.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; MAURANO, L. E. P.; NARVAES, I. S.; OLIVEIRA FILHO, F. J. B.; ROSA, R.; SILVA, E. B.; VALERIANO, D. M.; VICTORIA, D. C.; BAYMA, A. P.; SENTELHAS, P.C.; BATTISTI, R.; CÂMARA, G. M. S.; FARIAS, J. R. B.; HAMPF, A. C.; NEDEL, C. The soybean yield gap in Brazil – magnitude, causes and possible solutions for sustainable production. **Journal of Agricultural Science**, Toronto, v. 153, p. 1394-1411, 2015.

SENTELHAS, P.C., BATTISTI, R., MONTEIRO, L. A., DUARTE, Y, C. N., VISSÉS, F. de A. Yield Gap – Conceitos, Definições e Exemplos. **Informações Agrônômicas**, [S. l.], v. 155, p. 10 - 13, 2016.

SLC AGRÍCOLA. **Ciclo de Produção**. Porto Alegre: SLC Agrícola, [2022]. Disponível em: <https://www.slcagricola.com.br/modelo-de-producao/>. Acesso em: 10 out. 2022.

SOUSA, D. M. G; LOBATO, E. **Correção do solo e adubação da cultura da soja**. Planaltina: EMBRAPA, 1996.

TAGHIZADEH-MEHRJARDI, R.; NABIOLLAHI, K. A.; KERRY, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. **Geoderma**, Amsterdam, v. 266, p. 98-110, 2016.

TAGLIAPIETRA, E. L.; ZANON, A. J.; STRECK, N. A.; BALEST, D. S.; DA ROSA, S. L.; BEXAIRA, K. P.; RICHTER, G. L.; RIBAS, G. G.; DA SILVA, M. R. Biophysical and management factors causing yield gap in soybean in the subtropics of Brazil. **Agronomy Journal**, Madison, v. 113, n. 2, p. 1882–1894, 2021.

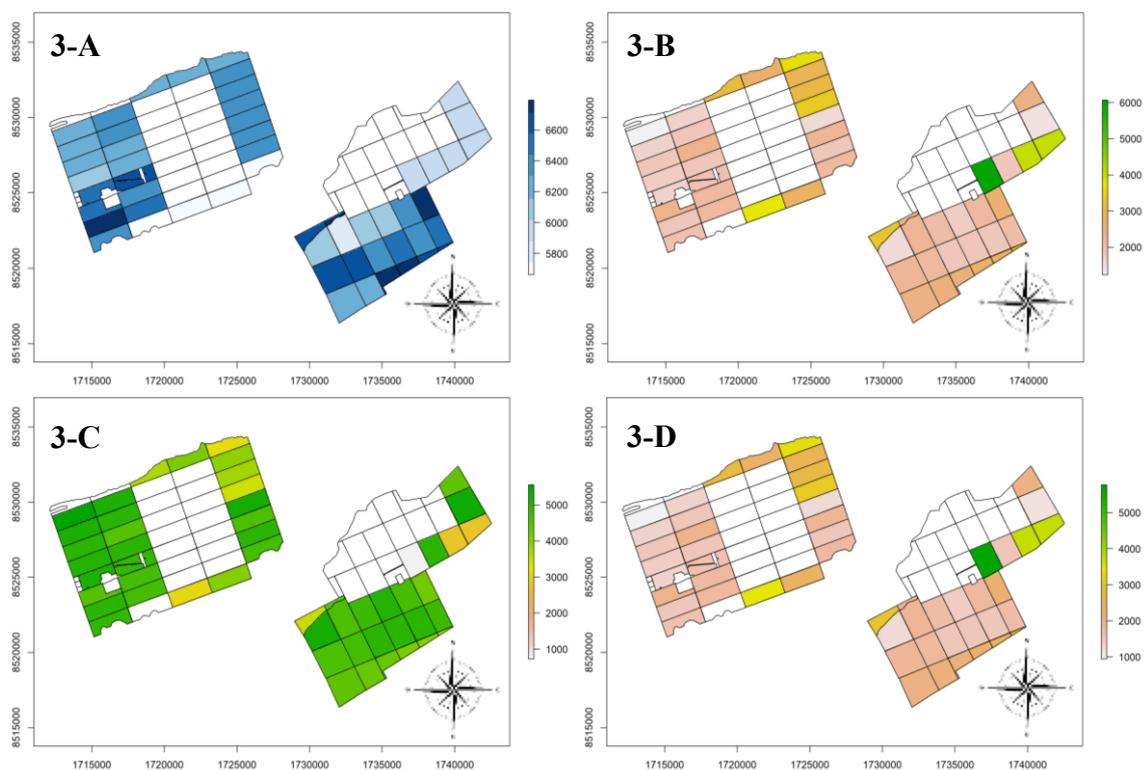
UNITED STATES DEPARTMENT OF AGRICULTURE – USDA. **Soybeans**: World Supply and Distribution. Washington, DC: USDA, [2020]. Disponível em: <https://www.fas.usda.gov>. Acesso em: 27 Sep.2020.

VAN RAIJ, B.; DE ANDRADE, J. C.; CANTARELLA, H.; QUAGGIO, J. A. **Análise química para avaliação da fertilidade de solos tropicais**. Campinas, SP: Instituto Agrônômico, 2001.

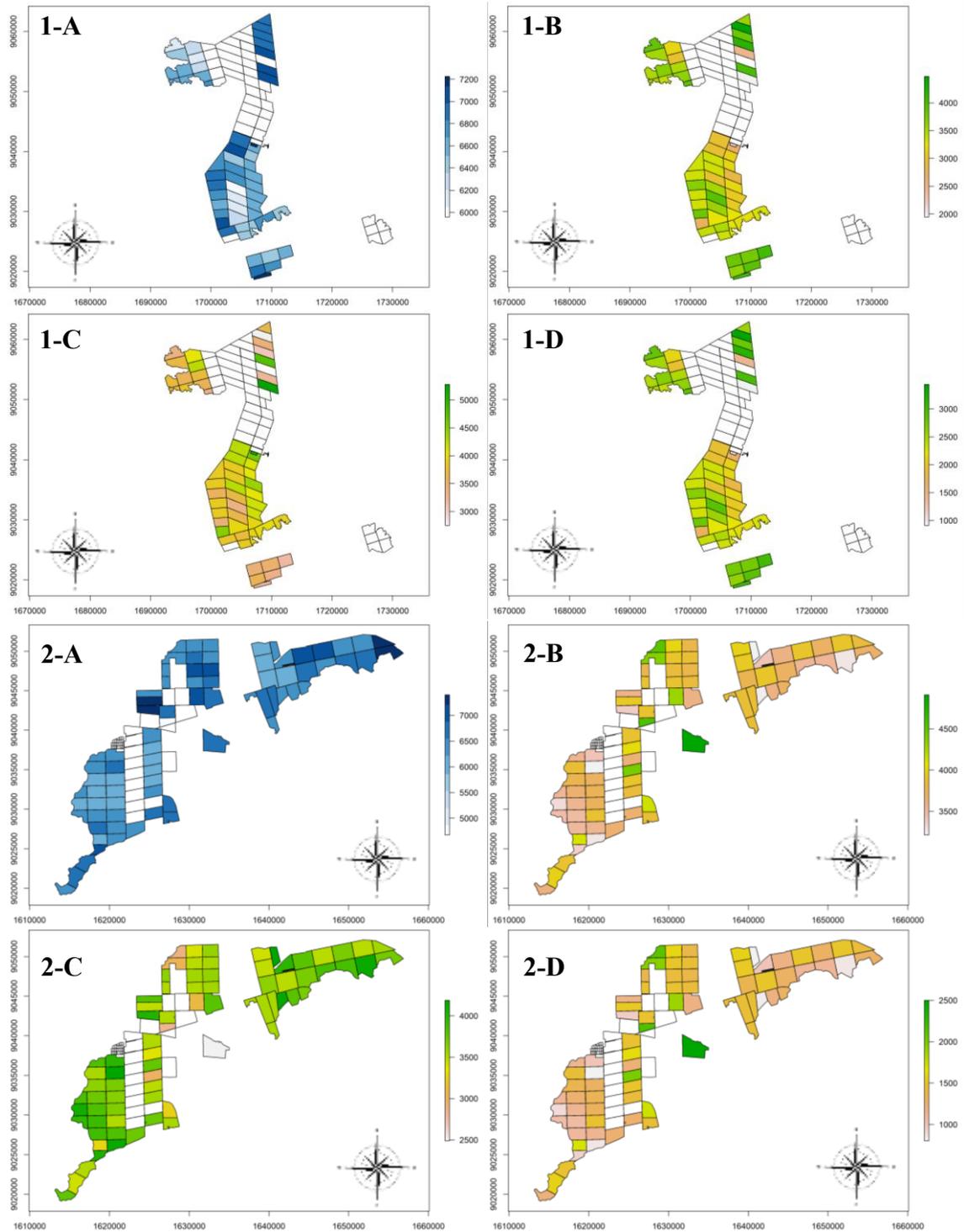
WALKLEY, A.; BLACK, I. A. An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. **Soil Science**, New Brunswick, N.J, v. 37, p. 29-38, 1934.

WILSON, D. R.; JAMIESON, P. D. Models of growth and wateruse of wheat in New Zealand. *In*: DAY, W.; ATKIN, R. K. (ed.). **Wheat Growth and Modelling**. Nova York: Plenum Press, 1984. p. 211-216.

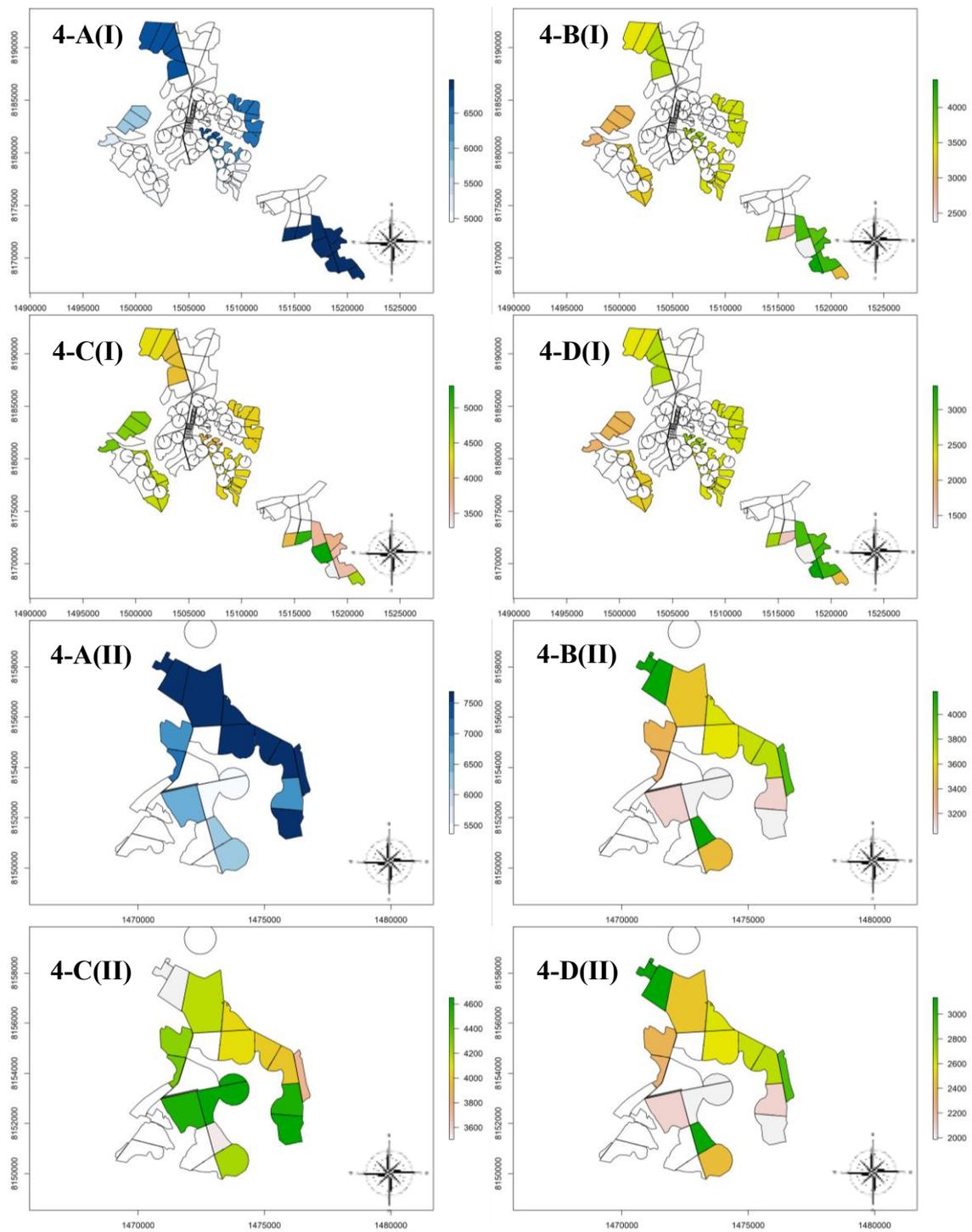
7. APÊNDICES



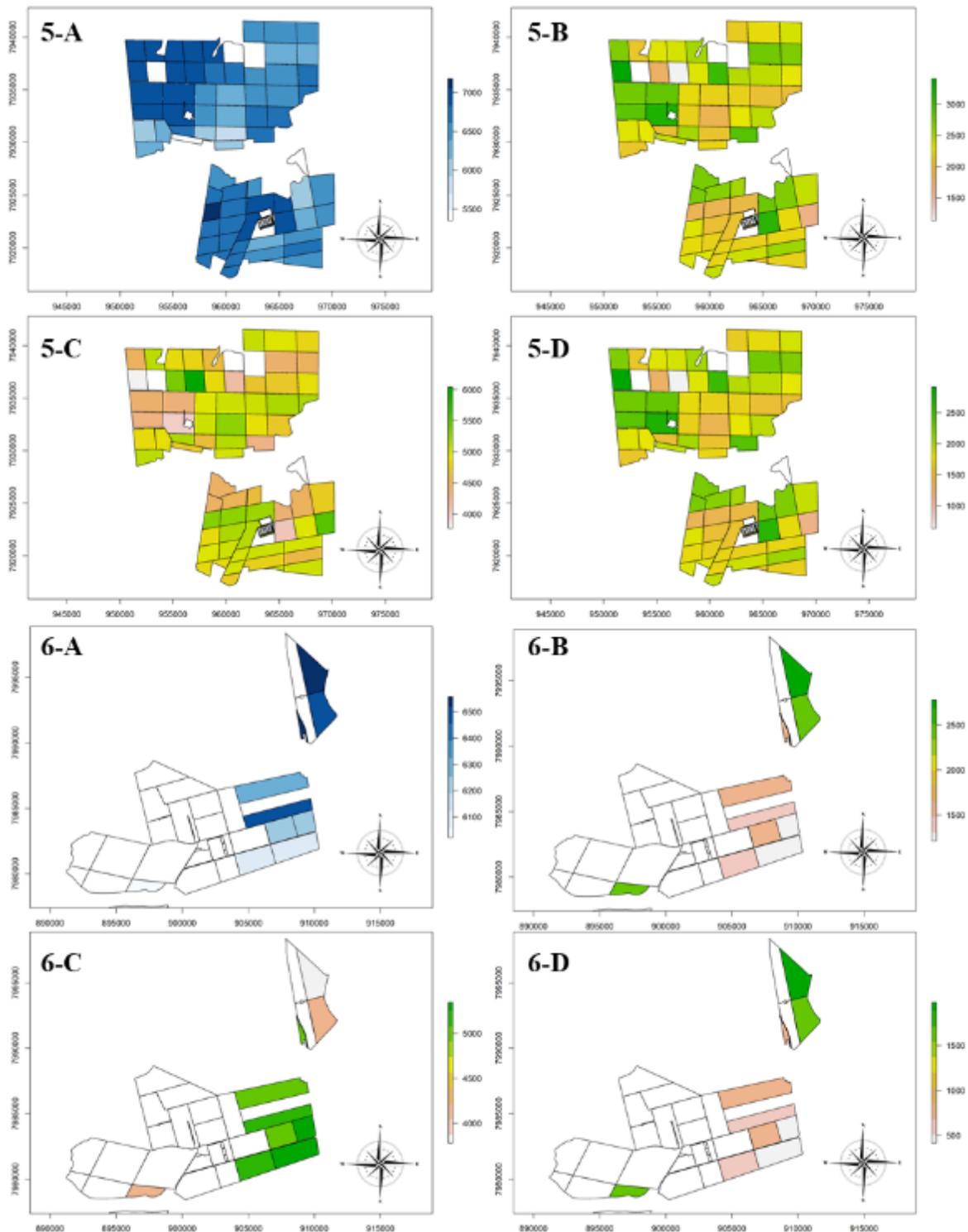
APÊNDICE 1. Mapas de classificação por talhão dos valores de (A), produtividade potencial (kg ha⁻¹), (B)YieldGAP total (kg ha⁻¹), (C)produtividade de soja obtida na safra 2021/22 (kg ha⁻¹) e (D)YieldGAP de manejo (kg ha⁻¹). Identificação: 3 – Fazenda Paladino.



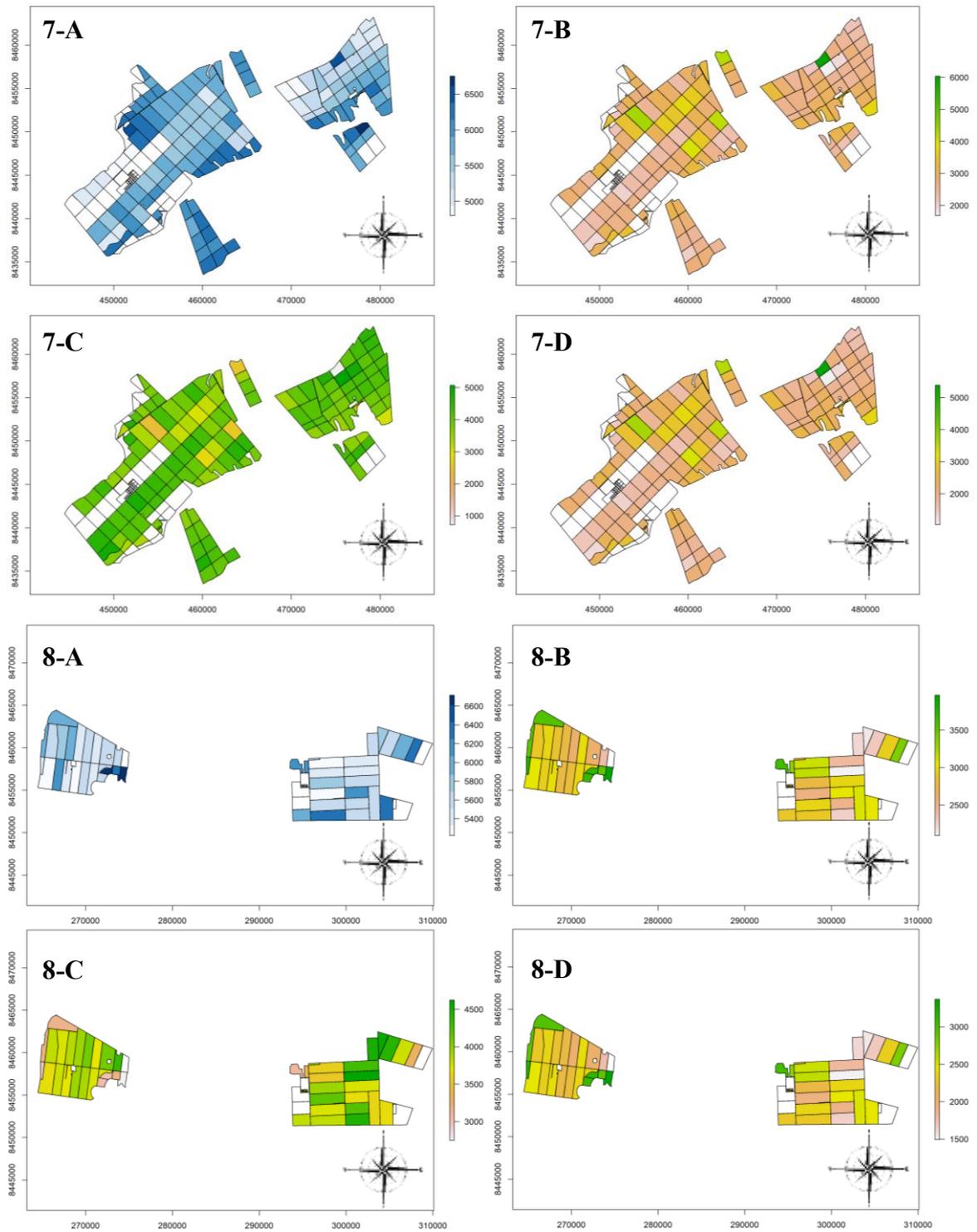
APÊNDICE 2. Mapas de classificação por talhão dos valores de (A) produtividade potencial (kg ha^{-1}), (B) YieldGAP total (kg ha^{-1}), (C) produtividade de soja obtida na safra 2021/22 (kg ha^{-1}) e (D) YieldGAP de manejo (kg ha^{-1}) Identificação: 1: Fazenda Parnaíba; 2 - Fazenda Planeste.



APÊNDICE 3. Mapas de classificação por talhão dos valores de (A)produtividade potencial (kg ha^{-1}), (B)YieldGAP total (kg ha^{-1}), (C)produtividade de soja obtida na safra 2021/22 (kg ha^{-1}) e (D)YieldGAP de manejo (kg ha^{-1}) Identificação: 4 – Fazenda Pamplona; (I) - Sede 1; (II) – Sede 2.



APÊNDICE 4. Mapas de classificação por talhão dos valores de (A)produtividade potencial (kg ha^{-1}), (B)YieldGAP total (kg ha^{-1}), (C)produtividade de soja obtida na safra 2021/22 (kg ha^{-1}) e (D)YieldGAP de manejo (kg ha^{-1}). Identificação: 5 – Fazenda Pantanal; 6 – Fazenda Planalto.



APÊNDICE 5. Mapas de classificação por talhão dos valores de (A)produtividade potencial (kg ha^{-1}), (B)YieldGAP total (kg ha^{-1}), (C)produtividade de soja obtida na safra 2021/22 (kg ha^{-1}) e (D)YieldGAP de manejo (kg ha^{-1}). Identificação: 7 – Fazenda Paiaguás; 8 – Fazenda Planorte.